



הטכניון – מכון טכנולוגי לישראל
Technion – Israel Institute of Technology

ספריות הטכניון
The Technion Libraries

בית הספר ללימודי מוסמכים ע"ש ארווין וג'ואן ג'ייקובס
Irwin and Joan Jacobs Graduate School



All rights reserved

*This work, in whole or in part, may not be copied (in any media), printed, translated, stored in a retrieval system, transmitted via the internet or other electronic means, except for "fair use" of brief quotations for academic instruction, criticism, or research purposes only.
Commercial use of this material is completely prohibited.*



כל הזכויות שמורות

אין להעתיק (במדיה כלשהי), להדפיס, לתרגם, לאחסן במאגר מידע, להפיץ באינטרנט, חיבור זה או כל חלק ממנו, למעט "שימוש הוגן" בקטעים קצרים מן החיבור למטרות לימוד, הוראה, ביקורת או מחקר. שימוש מסחרי בחומר הכלול בחיבור זה אסור בהחלט.

**Relationships between environmental parameters and species composition
patterns and their implications for biodiversity surrogates – a multiple scale study**

Research Thesis

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Rafi Kent

Submitted to the Senate of
the Technion - Israel Institute of Technology

Ttevet 5771

Haifa

June 2010

The Research Was Done Under The Supervision of Prof. Yohay Carmel in the Faculty
of Civil and Environmental Engineering.

The generous financial help of the Technion is gratefully acknowledged.

CONTENTS

TERMINOLOGY	3
SYMBOLS AND ABBREVIATIONS	3
INTRODUCTION.....	5
BACKGROUND AND GENERAL OVERVIEW	5
MAJOR MOTIVATION FOR THE PROJECT	6
SPATIAL SCALE IN BIODIVERSITY STUDIES.....	9
IMPLICATIONS FOR BIODIVERSITY CONSERVATION	9
DATA AVAILABILITY	11
RESEARCH OBJECTIVES	11
OBJECTIVES	11
 CHAPTER 1 - PRESENCE-ONLY VS. PRESENCE-ABSENCE DATA IN SPECIES	
COMPOSITION DETERMINANT ANALYSES.....	12
INTRODUCTION.....	13
METHODS	15
<i>Presence-absence and presence-only datasets</i>	<i>16</i>
<i>Data type effect</i>	<i>18</i>
<i>Bias analysis.....</i>	<i>19</i>
RESULTS	19
<i>Environmental determinants of species composition.....</i>	<i>19</i>
<i>Bias analysis.....</i>	<i>21</i>
DISCUSSION	22

CHAPTER 2 - RELATIONSHIPS BETWEEN MAMMAL SPECIES COMPOSITION AND ENVIRONMENTAL PARAMETERS AT VARYING SPATIAL SCALES IN THE CONTIGUOUS USA	25
INTRODUCTION.....	25
RESEARCH HYPOTHESES.....	28
METHODS.....	29
RESULTS AND DISCUSSION.....	29
CHAPTER 3 - COMPARISON OF CLASSIFICATION METHODS FOR PRODUCING SURROGATES FOR BIODIVERSITY	35
INTRODUCTION.....	35
METHODS	38
<i>Study area</i>	38
<i>Species distribution data</i>	41
<i>Environmental data</i>	42
<i>Surrogate production</i>	42
<i>Biological surrogates</i>	43
<i>Environmental surrogates</i>	43
<i>Performance evaluation</i>	44
RESULTS	45
<i>Species representativeness</i>	45
<i>Domains evenness</i>	47
DISCUSSION	48

CHAPTER 4 - DEVELOPING BIO-ENVIRONMENTAL SURROGATES FOR BIODIVERSITY: A CASE STUDY FROM MT. CARMEL, NORTHERN ISRAEL	50
INTRODUCTION.....	50
<i>Study area</i>	53
<i>Species distribution data</i>	53
<i>Production of surrogates for biodiversity</i>	55
<i>Biological surrogates</i>	55
<i>Bio-environmental surrogates (BES)</i>	56
<i>Performance evaluation</i>	57
<i>Surrogate fragmentation</i>	57
RESULTS	58
<i>Surrogate representativeness</i>	58
<i>Surrogates fragmentation</i>	58
DISCUSSION	63
SYNTHESIS	65
<i>Summary and conclusions</i>	67
APPENDIX 1 – A LIST OF WOODY AND GEOPHYTE PLANT SPECIES IN THE CARMEL SURROGATE ANALYSES	83

List of tables

Table i: Commonly used terms used throughout the dissertation – page 3

Table 2.1: A list of the different sampling units used in the analyses – page 27

Table 2.2: Environmental parameters used in the analyses – page 31

Table 3.1 - Environmental parameters used for logistic regression models and environmental surrogates for biodiversity – page 40

Table 3.2 - Summary of the rankings of all different combinations of number of classes and clustering algorithms – page 45

Table 4.1: Environmental parameters used in the study for logistic regression models and environmental surrogates for biodiversity – page 53

Table 4.2: The different types of surrogates for biodiversity – page 55

List of figures

- Figure i: A schematic flow-chart illustrating the parallel paths of the different parts of the project – page 6
- Figure 1.1: Distribution maps of two of the virtual species – page 17
- Figure 1.2: Three of the ordination diagrams – page 20
- Figure C :1.3CA λ values of the different environmental parameters – page 21
- Figure 1.4: Difference, in percent, between the average values of the different environmental parameters in locations with observations and the entire study area (contiguous USA) - page 22
- Figure 2.1: A map of the study area (the contiguous USA) – page 27
- Figure 2.2: The amount of explained variance in the composition of mammal species in the contiguous USA – page 30
- Figure 2.3: Amount of explained variance at a spatial scale of grain size and extent of 4,000 and $3.6 \times 10^6 \text{ km}^2$ respectively – page 32
- Figure 2.4: Amount of the variance in species composition explained by individual parameters – page 34
- Figure 3.1: Biological surrogates with eight classes, based on woody species distribution, produced with five different clustering algorithms – page 39
- Figure 3.2: A map of the study site in northern Israel – page 40
- Figure 3.3: Average number of species represented by a set of surrogates for the eight classes scheme – page 46
- Figure 3.4: Simpson's diversity index for each type of surrogate – page 47
- Figure 4.1: The average number of geophytes (top) and woody species (bottom) represented by each surrogate – page 59
- Figure 4.2: Comparative maps of surrogates for biodiversity on Mt. Carmel - page 60

Figure 4.3: Measures of fragmentation of the different surrogate types – page 63

Figure 5.1: Relative importance of the different parameter classes – page 67

ABSTRACT

The relationships between species and their environment is a central issue in ecology. Understanding these relationships can contribute to the conservation of biodiversity. However, there is a lack of data to conduct large scale species composition analyses, needed for large scale conservation planning. This work attempts to promote the scientific tools of biodiversity conservation along two lines. First, methods to produce biological and environmental surrogates for biodiversity were compared, and a new measure of surrogate efficiency was proposed. Results indicated that the most efficient classification method for biodiversity surrogates production is Ward's method of minimum variance. In addition, a new type of surrogates was developed, the bio-environmental surrogates, combining biological and environmental data to improve the representativeness of surrogates. The new hybrid surrogates represented almost all of the target species, regardless of the specific approach taken to produce them. The second principal objective of this work was to quantify the variation in the relationships between species composition and its environmental determinants as a function of spatial scale. The main problem was data availability. At large scales, the only available data is presence-only data, which is considered inappropriate for such analyses. Thus, a simulation study was conducted, with virtual species based on real environmental data and true species occurrence data from the Global Biodiversity Information Facility, to test whether presence-only data are sufficient for large scale species composition analyses. The results of the simulation tests allowed me to commence with the final step of this work, which was analyses of the environmental determinants of terrestrial mammal composition in the contiguous USA, at varying spatial scales. The results of these analyses revealed that climatic parameters such as temperature and precipitation are the most important environmental variables determining species composition. Land-use – land-cover parameters were also relatively influential at small and intermediate scales; however their effect decreased at larger scales. Topography and primary productivity were of lesser importance at the

scales studied, and their effect further decreased with increasing scale. The two lines of progress achieved by this work may be combined to enhance the abilities of conservation scientists and practitioners to protect biodiversity at large scales. In light of the extremely high species extinction rate, estimated at $\sim 100 - 1000$ times higher than the background rate, and in light of latest climatic changes and habitat loss through land-use alterations both processes directly or indirectly related to human activity in past century, urgent actions are needed in order to protect the remaining biodiversity. This work contributes to these efforts in a substantial manner.

Terminology

Throughout the dissertation I use several terms describing different spatial and other elements. These terms are listed in the table below and the list of symbols and abbreviations.

Table i: Commonly used terms used throughout the dissertation

<i>Term</i>	<i>Description</i>
Domain	An area within the study area consisting of sites with similar characteristics
Grid-cell	The basic unit in a grid dividing the study area into equal-size units, equivalent to the grain size of the analysis
Grain size	The resolution of the analysis. Determined by the size of grid-cells
Extent	The geographical size of the area of interest, or the entire grid
Site	A single grid-cell within a domain
Representativeness	The efficiency of surrogates in representing target species within the study area
Evenness	The level of similarity in size between different domains of a specific surrogate map

Symbols and abbreviations

P-O – Presence-only

P-A – Presence-absence

BES – Bio-environmental surrogates

NDVI – Normalized Difference Vegetation Index

DEM – Digital Elevation Model

GIS – Geographic Information System

mdt1 – Mean Daily Temperature in the Coldest Month

mdt8 – Mean Daily Temperature in the Hottest Month

CCA – Canonical Correspondence Analysis

DTR – Distance to Nearest Road

DTU – Distance to Nearest Urban area

GBIF – Global Biodiversity Information Facility

λ – The proportional contribution of a variable to the eigenvalue of the first axis of a canonical analysis

alt – Altitude

Temp – Temperature

TempS – Temperature Seasonality

Prec – Precipitation

PrecS – Precipitation Seasonality

MaxT – Maximum Temperature

MinT – Minimum Temperature

LULC – Land-Use/Land-Cover

INTRODUCTION

Background and general overview

The relationships between species and their environment stand at the core of the science of ecology. Although it is considered a relatively new science, ecological observations date back to the Greek philosophers. However, the first modern ecologists were Darwin and Wallace, who were the first to relate spatial patterns of species to their physical environment (Darwin 1859 chapter 3, p. 86-110, Wallace 1891, chapter 1, p. 3-19). This work attempts to address the question of biodiversity distribution determinants in two manners, a basic theoretical analysis of multiple scale species – environment relationships; and an applicative investigation of surrogates for biodiversity and their efficiency as biodiversity conservation tools (Fig. i). Understanding the relationships between species composition and environmental variables may considerably improve the ability to conserve biodiversity, which is important in light of the species extinction crisis. Chapter two presents a multiscale quantification of the relationships between species composition and environmental conditions. Although the concept of relating environmental conditions to species composition is not new (Whittaker 1956, Pianka 1966, MacArthur 1972). Such analyses require data on the distribution of the entire species pool at very large spatial extents. However, data for such analyses are only available from biological collections, in the form of presence-only (P-O) data. P-O data are considered inappropriate for species composition analyses (See 'data availability' section and chapter one for a detailed explanation). In order to examine whether the data collection method (i.e. presence-only versus presence-absence) has a significant effect on the results of species-environment analyses, at large spatial scales a simulation study was conducted prior to analyzing actual species data. These analyses are presented in chapter one.

The second path of investigation in this work was the efficiency of biodiversity surrogacy. The importance of surrogacy stems from a lack of distributional data for most elements of biodiversity, and in most regions of the world. Chapters 3 and 4 deal directly with methods to improve the production process of surrogates for biodiversity, by means of comparison between common methods to produce surrogates for biodiversity, and proposing an extensive approach to evaluating the efficiency of the resulting surrogate maps in representing biodiversity. In addition a novel approach to producing surrogates is proposed, based on both biological and environmental data, combined into bio-environmental surrogates.

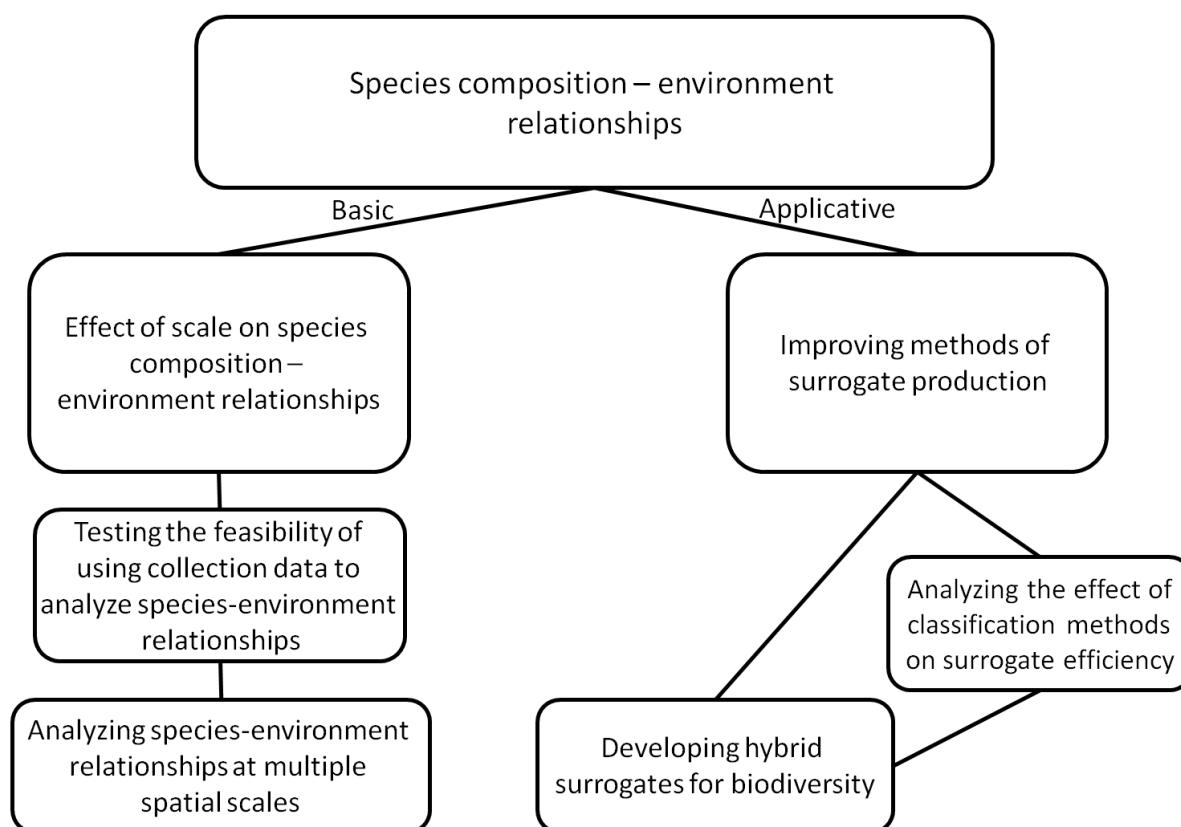


Figure i: A schematic flow-chart illustrating the paths of the different parts of the project.

Major motivation for the study

The rate by which species are going extinct is constantly increasing since the time of the industrial revolution, and is now approximately 1,000 times higher than the background extinction rate

(Ricketts et al. 2005). Most of this loss is attributed to human induced alterations of the environment, such as habitat loss, air, water and soil pollution and direct exploitation (Ehrlich and Wilson 1991, Shriner et al. 2006). Such high rates of extinction might have devastating effects on the well being of the entire planet, humans included. The disappearance of so many species may affect the functioning of natural ecosystems, to the level that ecosystem services – all the 'goods' provided to humans by various elements of the ecosystem – may be hampered (Worm et al. 2006). Some ecosystem services have obvious economic value, such as foods, drugs or construction materials. Less conspicuous services like reducing greenhouse gas concentrations in the atmosphere by carbon sequestration, drinking water purification by seepage, etc (Costanza et al. 1997) are crucial, but quantifying their value in monetary terms is not straight forward. Extinction rates have risen so high, that there is a dire need for actions to bring them to a halt, or at least slow them down substantially.

The aim of this study is to promote the understanding of processes in fields that might help in the slowing down of species extinction rates directly, by improving the efficiency of biodiversity surrogates as conservation tools (chapters 3 and 4) and indirectly, by improving the understanding of the relationships between species and their environment, and exploring additional information sources towards that end (chapters 1 and 2).

Species-environment relationships

Over the last 50-60 years there is continuous effort to come up with a single hypothesis explaining all observed patterns in the distribution of biodiversity, whether single species or entire ecosystems (e.g. Whittaker 1956, 1967, MacArthur 1972). The most obvious pattern is the latitudinal gradient in species richness, driven by varying of resource availability levels, of resources such as energy availability, evapotranspiration potentials and geometric constraints (limited space) (Pianka 1966, Crawley and Harral 2001, Whittaker et al. 2001, Allen et al. 2002, Algar et al. 2007). Species composition, an element of biodiversity that contains more information than species richness,

however is more difficult to explore, due to a lack of means for quantitative representation, was less frequently studied (Ferrier and Guisan 2006). Theories concerning patterns of species composition mainly include three conceptual models. A neutral model suggests that all the species are environmentally and competitively equal, and patterns are caused by random dispersal events (Hubbell 1997, 2005). Hubbell's theory is restricted to communities consisting of species in a single trophic level, and asserts that all the species in a given community have the same probability to inhabit free space created by disturbance, thus the only factors responsible for species assembly in the community are stochastic processes affecting propagule dispersal. Unlike the theory on island biogeography (MacArthur and Wilson 1967), which is also a neutral theory of species assembly, based on the assumption that each species has the same probability to inhabit an island, and that species accumulation is governed by the size of the island and its distance from the mainland, Hubbell's theory assumes that each individual has the same probability for settling in a new habitat patch (Hubbell 1997). The environmental model relates species composition to environmental conditions (e.g. Whittaker 1956, 1967). This model is based on the ecological niche theory. Each species reacts to its environment according to its physiological and ecological requirements. These relationships dictate the species' extent of occurrence. Thus environmental conditions dictate the presence of species in a given location. Whittaker (1956) concluded that most of the variability in tree species composition on the Great Smoky Mountains can be explained by two environmental gradients. According to the third model, species composition is determined by inter-specific interactions within and among trophic levels, through mechanisms of competitive displacement, local extinctions caused by predation etc. (May 1984). These models are non-mutually exclusive (Legendre et al. 2005). Chapter 2 of this study concentrates on the environmental model, attempting to understand which environmental factors are important at which spatial scale.

Spatial scale in biodiversity studies

The question of scale has become a key issue in ecological research. Since different processes may play a role at different scales, and since different organisms perceive the environment at different scales (Levin 2000), single-scale analyses may be incomplete. However, most studies on patterns of species distributions, richness and composition are still carried out at a single spatial scale. It is now well accepted that there is no 'right' scale for biodiversity analyses, regardless of organization level (Willis and Whittaker 2002, Wagner and Fortin 2005). Studies that were conducted at multiple scales either define scale qualitatively (Cushman and McGarigal 2002) or change only grain or extent of the analysis. However, the behavior of scale is not intuitive (Wiens 1989), since it is characterized by both grain and extent; modification of each component alone while keeping the other constant is not a complete change of scale, and thus may yield misleading conclusions, due to both principal and statistical difficulties. A change of grain alone does not constitute a full change of scale, rather it alters the resolution of the analysis within a single scale, because both the biotic and a-biotic components of the studied system remains unaltered, thus it affects the amount of detail in the analysis of the reaction of species to the analyzed gradients, without altering the length of these gradients. A change of extent without a change of grain only affects the length of environmental gradients observed, while pointing out the same environmental processes.

Implications for biodiversity conservation

Methods used in decision making for biodiversity conservation worldwide changed considerably over the years (Poiani et al. 2000). In the past, decisions were taken *ad hoc* and were aimed at preserving areas that were not used for other (profitable) purposes and which contained high biodiversity and/or 'highly popular' species, in order to attract tourism (Maddock and Du Plessis 1999, Prendergast et al. 1999). During the past few decades, focus shifted towards conservation of endangered species, and reserves were selected according to the distribution of those species (Stoms 2000). Recently, a more

'holistic' approach, also called 'coarse filter', has become more widespread. According to the 'coarse filter' approach, entire ecosystems and landscape units should be the aim of conservation, rather than single species (Margules and Stein 1989, Maddock and Du Plessis 1999, Schwartz 1999, Bonn and Gaston 2005). In addition, it is necessary to conserve a variety of habitats and ecological processes which facilitate the persistence of a myriad of species (Noss 1987, Pressey et al. 1993, Stoms 2000). Applying such approaches in conservation decision making requires knowledge of the distribution ranges of all species. Nevertheless, in most regions of the world species distributions data are partial at best, neither is it feasible to collect such data for all species. One way to deal with that problem is by using habitat-suitability modeling (Guisan and Zimmermann 2000). Habitat-suitability models rely on Hutchinson's niche theory (Hutchinson 1957, Vandermeer 1972), relating distributional data to environmental parameters in surveyed areas, defining the species' niche. By finding areas containing similar conditions outside the surveyed area, it is possible to estimate the presence or absence of the species in un-surveyed areas. These models include General Linear Models, General Additive Models and machine learning models such as Neural Networks. However, models that predict species presence in un-surveyed areas require dedicated survey effort for each species individually in a confounded area, which is both time consuming and expensive. In addition, they are imperfect, and when dealing with a multitude of species, uncertainty accumulates and is very difficult to incorporate into the analysis reliably. Thus, it is widely accepted to use surrogates for biodiversity (Belbin 1993, 1995, Ferrier and Watson 1997, Fairbanks 2000, Carmel and Stoller-Cavari 2006, Sarkar et al. 2006) which is a less expensive method, and is probably also less prone to accumulating uncertainty. Although most conservation planning and actions are conducted at relatively local scales of 10^2 - 10^3 km, most studies on surrogate efficiency take place at much larger scales (10^4 - 10^6 km) (Carmel and Stoller-Cavari 2006, Rodrigues and Brooks 2007). Such studies rely on either range maps from atlases or low resolution environmental data (see review by Rodrigues and Brook 2007). While

results might indicate strong congruence in patterns over such large spatial extents, they might not be consistent at extents relevant for conservation. Studies conducted at relatively small spatial scales usually indicate that both types of surrogates perform better than random choice of conservation areas (Sauberer et al. 2004, Trakhtenbrot and Kadmon 2005, Carmel and Stoller-Cavari 2006, Trakhtenbrot and Kadmon 2006),

Data availability

A common thread in every part of this work is the lack of reliable data on the distributions of the analyzed species. The study of species composition – environment relationships requires spatial data on large numbers of species, and for large geographical extents. Such data are scarce, especially in the most diverse areas of the world (Elith et al. 2006, Ferrier and Guisan 2006, Loiselle et al. 2008). On the other hand, availability of occurrence records via internet is rapidly growing (e.g. the Global Biodiversity Information Facility, GBIF 2008). Occurrence data are also termed presence-only data, since they lack explicit information on unvisited locations, and are thus considered inappropriate for analyzing species-environment relationships. There is a large body of literature regarding the use of presence-only data in species distribution modeling, and in species richness patterns, but with no conclusive evidence (Brotons et al. 2004, Elith et al. 2006, Tsoar et al. 2007).

Research objectives

This study is concerned with the relationships between species and environment, how these relationships are affected by spatial scale, and in turn, how they affect surrogates for biodiversity. Thus, the main research objectives and questions are:

Objectives

1. To combine environmental and biological surrogates into a new type of surrogates (Bio-Environmental Surrogates).

2. To evaluate their efficiency in representing biodiversity, and compare it to the efficiency of existing types of surrogates.

3. To understand the environmental determinants of biodiversity patterns at multiple spatial scales.

4. To quantify the variation in the effect of different environmental parameters on species composition at various spatial scales.

Ideally both path of the study would be carried out in the same geographic area. However, data limitations dictated the selection of study sites according to the amount of available data. In the contiguous USA there is a large amount of collection data, digitized and readily available via the internet. On Mt. Carmel, on the other hand, I had access to high resolution, high quality dataset, suitable for a surrogacy study at small spatial scales.

Chapter 1

Presence-only vs. Presence-absence data in species composition determinant analyses

INTRODUCTION

Studying the relationships between species and their physical environment requires data on the distribution of species in space. Ideally, such analyses would be based on presence-absence data (P-A), collected through dedicated surveys. However, such data are scarce, and exist only for areas of small spatial extent, and are especially uncommon in the most diverse areas of the planet (Elith et al. 2006, Ferrier and Guisan 2006, Loiselle et al. 2008, Sastre and Lobo 2009). Presence-only (P-O) data have various shortcomings in regards to analyses of species-environment relationships, for example (a) They lack explicit information on unvisited locations, and (b) they might contain errors and biases. Potential biases include spatial bias (concentration of observations in easily accessible locations and over sampling of species-rich areas, Ponder et al. 2001, Kadmon et al. 2004), taxonomic bias (over-representation of certain species, Hijmans et al. 2000), and environmental bias (under-representation of areas at the edges of the environmental gradient, Loiselle et al. 2008). In order to quantify the amount of bias in the data set, analyses were carried out to quantify the amount of environmental and geographical bias in a relatively large sample of observation locations in GBIF (GBIF data portal, <http://data.gbif.org>). In contrast to these shortcomings, such data are readily available in large quantities and, due to an accelerating effort to digitize and publicize these data, are also highly accessible (Graham et al. 2004). The validity of using presence-only data in ecological analyses was studied several times. Results have been inconclusive, with some authors reporting sufficiency of presence-only data (Elith, Graham et al. 2006; Loiselle, Jorgensen et al. 2008), superiority of presence-absence data (Guisan and Zimmermann 2000; Hirzel, Helfer et al. 2001; Graham and Hijmans 2006), or differential success for different species (Elith et al. 2006,

Tsoar et al. 2007). The focus of most studies was the distribution of a single species or modeling species richness patterns. Models that utilize P-O data to predict species distributions in unsurveyed areas include ENFA (Ecological Niche Factor Analysis) which compares the distributions of environmental factors (termed EcoGeographical Variables) (i.e. mean and variance) in locations where the species was observed with the distributions of the same EGVs in the entire study area (Hirzel et al. 2002). A different model (MaxEnt) estimates the potential distribution of species based on machine learning approach (Phillips et al. 2006).

Ferrier and Guisan (2006) reviewed approaches to community-level modeling. They used both P-A P-O data for their models, and concluded that P-O data are problematic for such analyses. They stated that due to data limitations, analyses of species composition are limited to areas of small spatial extent. To the best of our knowledge, the value of P-O data for studying species composition determinants at large spatial scales was seldom evaluated before (but see Kadmon and Heller 1998, Yom-Tov and Kadmon 1998, Kadmon and Danin 1999).

In this study I attempt to determine whether, when considering a multitude of species over a large spatial extent, data type (P-A vs. P-O) has a significant effect on the results of analyses of species-environment relationships. A direct comparison of the effect of data type on the results of such analyses requires complete datasets of the two types, containing data on the same species and with the same spatial extent. Therefore, available data of actual observations is not optimal for such analyses, and a simulation study seems to be the most appropriate solution. P-A and P-O datasets of virtual species within the contiguous USA were simulated and the effect of data type on the results of multivariate analyses of the environmental determinants of species composition was examined. In order to keep the simulations as close to reality as possible, real environmental data were used to define species niches, As well as real locations of observations to create a P-O sampling scheme, incorporating real biases into our datasets. The null hypothesis of this chapter was that the sampling

technique (i.e. P-A or P-O) will not affect the results of multivariate analyses of multiple species datasets covering a large spatial extent.

METHODS

Distribution maps for 50 virtual species within the land area of the contiguous USA were produced. Species distributions were based on niches reflecting actual environmental conditions in the study area. The realized ecological niche of each virtual species was defined by selecting a random location within the study area to represent the niche center in parametric space and recording the values of six environmental parameters in this location: max. temperature of the hottest month (MaxT); min. temperature of the coldest month (MinT); annual precipitation (Prec); altitude (alt); normalized difference vegetation index (NDVI); and distance to nearest urban area (dtu). Climatic and topographic parameters were taken from Worldclim (Hijmans et al. 2005). An NDVI layer was downloaded from MODIS (<http://glcf.umd.edu/data/ndvi>) and produced a layer of distance to nearest urban area from a map of the urban areas of the USA (data was extracted from ESRI data files (ESRI 1999)). The correlation level between each pair of variables at 200 random locations within the study area were tested. The average correlation level was 0.36, and the maximum correlation was 0.84. The marginal effect of each parameter on species composition, as analyzed by Canonical Correspondence Analysis (CCA) is the effect each variable has regardless of other variables, thus, these correlations are evident in the analyses, but do not affect the results (Ter Braak 1986). All environmental layers were rescaled to a resolution of 0.0833^0 (~10km). Niche breadth was set as a random fraction (between 0.05 and 0.5) of the true range of each parameter in the study area, above and below the niche center. Simulations were carried out in MatLab (MathWorks, Natick, Massachusetts, USA). Distribution maps were produced for each virtual species in ArcGIS (ESRI 1999) by superimposing a grid with mesh size of 0.0833^0 (~10km) over the entire study area.

Grid cells were assigned a value of 1 where all environmental parameters were within the specified realized niche, and zero otherwise (Fig. 1).

Presence-absence and presence-only datasets

To produce a P-A dataset 1,072 locations from the geographic space of the entire study area were selected (1,072 is the median number of sampling locations found in 17 studies that used presence-absence data). For each location the presence and absence of all virtual species was recorded, producing a matrix of 50 columns and 1,072 rows. The values of eight environmental parameters (the same six parameters used for defining niches, plus temperature seasonality (standard deviation*100; TempS) and precipitation seasonality (coefficient of variance; PrecS)) were recorded in the same 1,072 grid cells, resulting in an environmental data matrix of eight columns by 1,072 rows. The two additional parameters were expected to show a weaker relationship with species composition than the parameters used to define the niches. Since they are correlated to parameters defining the species' niches, some relation between them and the distribution of the species was expected.

P-O data typically contain spatial bias towards easily accessible locations, as well as areas with high biodiversity (Hijmans et al. 2000). In order to incorporate such bias into our datasets, the locations of real observations of a random selection of avian species in the contiguous USA were used, using GBIF (GBIF 2008). A compilation of ~200,000 observation locations was derived from real observations in the GBIF dataset, hereafter the observation pool. The distribution range of virtual species j contained a subset of N_j records from the observation pool. In order to mimic taxonomic bias, as it exists in observations of real species, 50 avian species we selected randomly, and the number of observations existing for them in GBIF was recorded. Each virtual species was randomly assigned a number of observations of one of the 50 avian species (n_j). For each species j , n_j observation locations were randomly selected, out of the N_j observations located within its

occurrence range. That resulted in a matrix of 50 columns, denoting the 50 virtual species and, initially, 120,670 rows (the number of grid cells in the entire study area). Next, all empty cells were deleted from the matrix (cells with no species present). Mean size of the full P-O sets was 24,696 observations. Since there was an order of magnitude difference between P-A and P-O dataset sizes, partial sets of P-O data were also produced, consisting of a random choice of 1,072 rows from the P-O matrix, resulting in P-O datasets of the same size as the P-A datasets. Sets of each data type (P-A, P-O and partial P-O) were produced five times, independently.



Figure 1.1: Distribution maps of two of the virtual species. One with a wide niche (top panel); and one with a narrow niche (bottom panel).

Data type effect

Canonical correspondence analysis (CCA) was used to examine whether data type affects the results of analyses of the relationships between species composition and environmental parameters (Ter Braak and Verdonschot 1995) using CANOCO 4.5 (Ter Braak and Smilauer 2002). CCA is an ordination technique that performs gradient analyses, constrained by species composition, iteratively (Ter Braak 1986, Legendre and Legendre 1998). Ordination is the simplification of a multi-dimensional space by reducing the number of axes in this space (Legendre and Legendre 1998). The reduction is achieved by extracting the major gradients from the explanatory variables, which explain the largest amount of the variance in the independent variable distribution, and creating axes that represent these gradients. CCA assumes that the relationships between environmental parameters and species composition are unimodal, rather than linear as do Principal Component Analysis and General Linear Models (Ter Braak 1986, Legendre and Legendre 1998).

CCA analyses were applied on the three dataset types, and compared the contribution of the various environmental parameters as explanatory variables determining species composition. Each CCA analysis resulted in a λ value for each parameter. $\lambda(x)$ is the proportional contribution of variable x to the eigenvalue of the first axis. Another element of the ordination is the relationships between the various parameters, i.e. the level of correlation and the directionality of their effect on species composition (Ter Braak 1986, Ter Braak and Verdonschot 1995). The ordination diagrams were examined in order to qualitatively explore the relationships among the different parameters and between them and the virtual species, within the ordination space.

Although CCA analyses are not normally repeated, and do not require replications, I repeated the analyses five times, to ensure the consistency of our results. Thus, univariate analysis of variance was carried out in SPSS using the different parameters as covariates and the different data types as fixed

factors, to determine whether the differences in the amount of variance explained by the environmental parameter λ (s values) obtained from P-A and P-O data, both full and partial.

Bias analysis

Analyses on the observation pool dataset and its respective environmental dataset were carried out in order to quantify the amount of bias in the data. Environmental bias was calculated as the difference between values of each environmental parameter in the entire study area (contiguous USA) derived from (1) all grid cells in the study area and (2) all the locations of the actual observations in the observation pool (~200,000). From the environmental bias, spatial bias (difference in distance to nearest urban area, as described above) was extracted.

RESULTS

Environmental determinants of species composition

Ordination diagrams produced for each data type revealed that environmental parameters had similar effect on species composition, in all datasets regardless of data type (Fig. 1.2). For example, distance to nearest urban area and altitude were highly correlated in their effect on species composition. The effect of data type on the amount of variance explained by each variable was examined. I expected that the effect of distance to nearest urban area (dtu) would be less prominent when using P-O data, since the range of values of this parameter was smaller in P-O data than in randomly selected locations (P-A). In contrast, the results showed that there was no significant difference between the amount of variance explained by this factor in the two data types (Kruskal-Wallis ANOVA, $\chi^2 = 4.455$, $p = 0.108$). Univariate ANOVA, with environmental parameters as covariates and data type as a fixed factor ($F=2.553$, $p=0.082$, Fig. 1.3) showed there was no significant effect of data type on the results of the CCA.

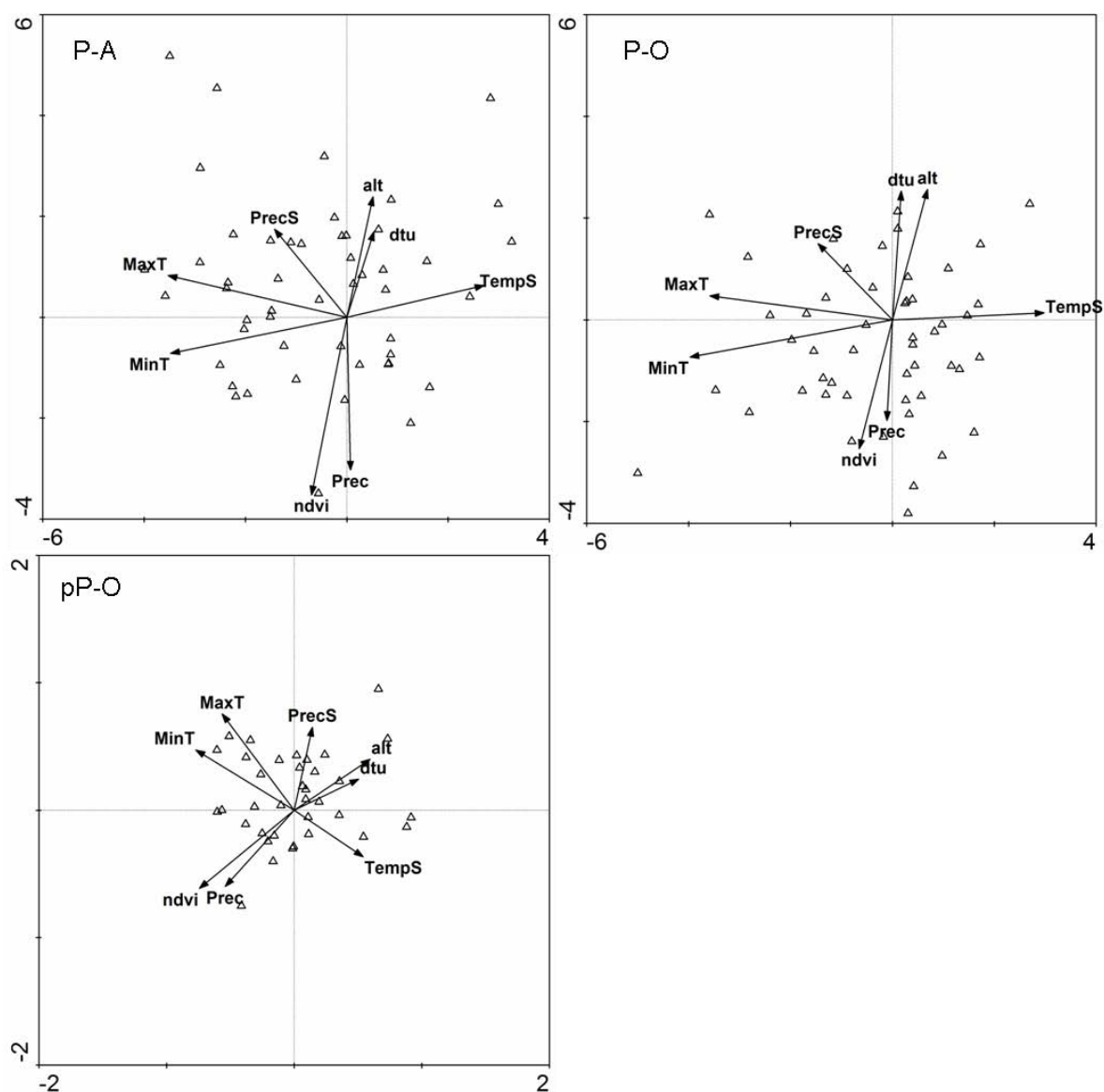


Figure 1.2: Three of the ordination diagrams. In the upper left corner is an ordination diagram of one CCA repetition applied to Presence-absence (P-A) data for 50 virtual species in 1072 sites. In the upper right corner is an ordination of one CCA repetition of Presence-only (P-O) data for 50 virtual species in ~25,000 sites. In the bottom left diagram is an ordination of partial P-O data (a subset of 1072 sites out of ~25,000 P-O data). Relationships between the various parameters (arrows) are similar in all diagrams, as well as the strength of their effects on species composition. Partial P-O is rotated around the origin of the axes, yet the relationships between the parameters and the species, as well as among the different parameters are similar to those in the P-A and P-O diagrams.

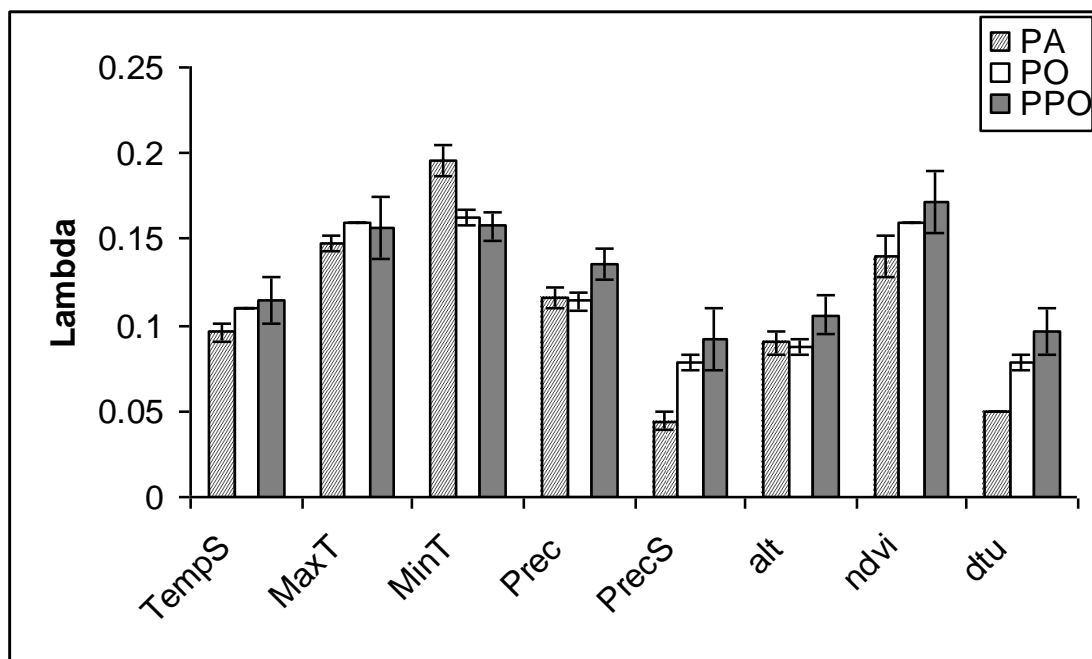


Figure 1.3: CCA λ values of the different environmental parameters are average values over five repetitions. Error bars are standard deviations. Univariate ANOVA showed no significant effect of data type on λ values when the different parameters were used as covariates and data type as a fixed factor (SPSS).

Bias analysis

Mean values of minimum temperature in the coldest month, distance to nearest urban area and altitude were lower in the locations of the observation pool than in the entire study area. In contrast, mean values of NDVI and annual precipitation were higher in these locations (Fig. 1.4). Ranges of all parameters were similar between the sampled area and the entire study area.

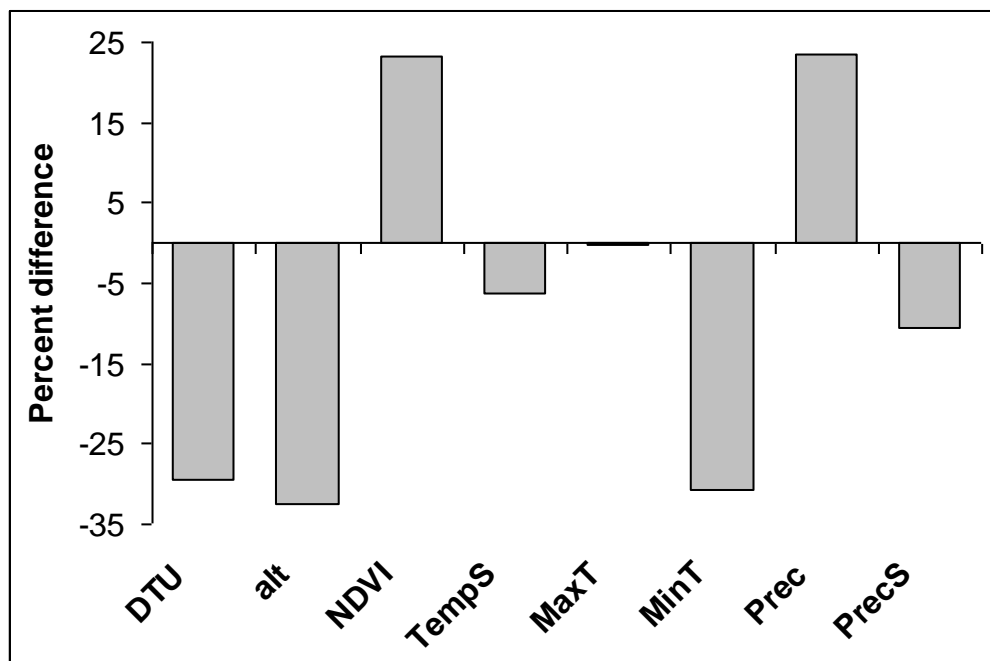


Figure 1.4: Difference, in percent, between the average values of the different environmental parameters in locations with observations and the entire study area (contiguous USA). Positive values denote higher sample values and negative values denote higher study area values.

DISCUSSION

Results support the research hypothesis and show quantitatively, for the first time, that P-O data can be used to characterize the relationships between environmental variables and species composition over large spatial extents. I found, by using virtual species for which complete distributional information is available, that CCA is robust enough to identify the main environmental drivers of species composition despite the bias contained in such data.

Results of the CCA analyses were highly consistent, showing similar effect of the various environmental parameters on species composition, regardless of data type. This consistency implies that the method is not sensitive to data type, and that the bias in the GBIF data does not significantly affect the outcome of analyses, at least at large geographical extents.

As expected, the results only partially explained the variance in the data, due to the non-unimodal relationships of the simulated species with environmental variables. However, it has been suggested that such simple representations of species should be used to test the robustness of ordination

techniques, such as CCA (Minchin 1987). The results suggest that CCA is robust enough to give consistent results despite the violation of the unimodal relationship assumption. partial P-O data were used in order to account for the difference in sample size between P-A and P-O (an order of magnitude). One might expect that the larger amount of data in the P-O set might compensate for its assumed relatively poor quality. Yet, partial P-O analyses results were very similar to those of P-O, suggesting that the amount of data had little effect on the results. All three datasets revealed similar relationships between environmental parameters and species composition.

Given the disagreement among authors regarding the value of P-O data for species distribution modeling, and the conclusions of Ferrier and Guisan (2006) that P-O data are insufficient for community-level modeling, the results presented here may seem surprising. One plausible explanation for my results is related to the amount of information within a dataset. Analyses of species-environment relationship require dividing the studied area into grid cells. Since typically most grid cells in a given study area are empty in P-O datasets (Ferrier and Guisan 2006), the amount of information in the occupied cells is crucial for the success of the analysis. When analyzing data from multiple species, each grid cell may contain data on more than one species. The cells may thus contain more information than in a single-species analysis. In addition, the number of occupied cells is dependent on the number of species, due to the larger amount of observations, as well as on the spatial distribution of the observations in the dataset. Thus, using multiple species increases the amount of data available for the interpretation of the species-environment relationships. This may explain the doubts regarding P-O data for individual species modeling, as well as my success in using P-O data for community level analyses. Fifty virtual species were used, which is a relatively small number of species, compared to actual species numbers found in such large areas, e.g. there are >400 mammal species (Kays and Wilson 2002) and >900 bird species (<http://www.birdlist.org/usa.htm>) in the contiguous USA. Thus, analyses based on real species may

be even more robust and reflect the true species-environment relationships. The effect of multiple species on the consistency of the results of the analyses is apparent in my results. There are differences in the locations of specific species in relation to the different environmental variables in the ordination space of the various data-types, suggesting that single species analyses may be more sensitive to data-type than multi-species analyses.

Kadmon et al. (2004) incorporated roadside bias correction when modeling species distribution with bioclimatic models. They concluded that such corrections should be incorporated only posteriori to an examination of the amount of environmental variability between near-road locations and off-road locations. They also suggested that in an area of small climatic variance between the road network and the entire area, roadside survey data are appropriate without correction. In a simulation study, Reese et al. (2005) found that using data that contain roadside bias may produce model results that do not differ much from models based on systematic surveys.

The analysis revealed that the observations in the GBIF database indeed included environmental and geographical biases. Observations were biased towards areas of high primary productivity, higher annual precipitation, higher minimum temperatures and lower altitudes. Seasonality had a small effect on observation frequency, probably due to the relatively low temporal resolution of the data. All the differences indicate that observers tended to look for species in productive areas, where conditions are relatively convenient, and avoid extreme environments. Geographical bias was represented by the average distance from the nearest urban area, under the assumption that observations will be concentrated closer to urban areas than would be expected by chance. The analysis indicated that the average distance of observations to an urban area was indeed ~30% smaller than the average distance in the study area, suggesting a strong bias towards sampling 'close to home'. In spite of these biases, the results were robust and consistent and the findings thus

confirm that easily accessible, web-based data are indeed amenable for the study of large-scale species composition determinants.

Chapter 2 - Relationships between mammal species composition and environmental parameters at varying spatial scales in the contiguous USA

Introduction

Understanding the factors affecting the distribution of biodiversity in time and space is a central objective of ecology (Shmida and Wilson 1985). Relationships between environmental parameters (e.g., climate, topography) and biodiversity patterns are scale-dependent both spatially and temporally (Levin 2000). Species richness, probably the most studied aspect of biodiversity, was often shown to vary as a function of spatial scale (Rahbek and Graves 2001, Whittaker et al. 2001, Nogués-Bravo et al. 2008).

Theories concerning the mechanisms governing distribution patterns of biodiversity elements, range from global (latitudinal species richness gradient) to very local scales, and relate to various processes, such as environmental, historical and evolutionary (Rosenzweig 1995). MacArthur (1972) presented ten alternative explanations for such patterns, related to historical events, inter-specific interactions and climatic conditions. He offered a general theory to explain patterns of species richness, i.e. the principle of equal opportunity. However, the mechanistic explanation of species richness patterns does not explain variation in species composition. Patrick (1963) found similar numbers of diatom species in different streams, while species identity differed considerably. Whittaker (1956) reported that most of the variation in tree species composition in the Great Smoky Mountains could be explained by two environmental gradients i.e. elevation and water balance. Recently it has been suggested that species richness patterns are largely determined by historical biogeographical processes (Pyrone and Burbrink 2009). Theories on species composition patterns include a neutral

model suggesting all differences are caused by random differences in dispersal between demographically and competitively equal species (Borcard et al. 1992, Hubbell 1997), an environmental model, relating species distributions to environmental conditions (Legendre et al. 2005), and a model suggesting that species composition is determined by interspecific interactions within and between trophic levels (May 1984). These models are non-mutually exclusive. From a historical perspective, in 1916 Clements (Clements 1916) suggested that communities (i.e. species assemblages) are discrete, and there are stable associations between species, supporting the interspecific interactions theory. On the other hand, Gleason (1939) asserted that individual species react to their physical environment, thus there is gradual change in species composition along environmental gradient. Species composition was only seldom studied at multiple spatial scales, but see for example Grand and Cushman (2003) and Grand and Mello (2004). In these studies, scale was defined qualitatively, i.e. plot, patch and landscape scale. However, most of the studies on species composition were restricted to a single scale (Svenning and Skov 2005, Rodriguez et al. 2006, Jones et al. 2008). Applied across multiple scales, multivariate analyses may provide a wider picture of the relationships between environmental parameters and species composition (Ter Braak 1986, Cushman and McGarigal 2002). Understanding species composition – environment relationships, and specifically how they are affected by spatial scale, improves conservationists' ability to predict the spatial distribution of biodiversity entities, and their reaction to global and regional changes (Margules and Pressey 2000).

There are serious conceptual and practical impediments to such analyses. A central conceptual challenge is the nature of scale, whose behavior is not intuitive (Allen and Hoekstra 1992). Scale is characterized by both grain (grid cell size) and extent (Willig et al. 2003). In most studies, a change of only a single element of scale is regarded as a full change of scale (Wiens 1989). Here a 'complete' approach was used, in which both grain and extent are modified together in the process of upscaling.

Four scales were used, with grains ranging between 100 and 10,000 km², and extents between 90,000 and ~10,000,000 km². The lower limit of spatial scale was the amount of data in GBIF. At smaller scales the amount of occurrence data on mammals is insufficient for ordination analyses. At each spatial scale a number of spatial units were sampled and analyzed independently (Table 2.1 and Fig. 2.1).

Table 2.1: A list of the different sampling units used in the analyses including: Grain size and extent (in km²), location, number of occurrence records and number of species.

Unit	Grain size	Extent	Location	No. of occurrences	No. of species
1	100	9*10 ⁴	West (North)	2,510	56
2	100	9*10 ⁴	West (South)	870	51
3	100	9*10 ⁴	East (North)	1351	37
4	100	9*10 ⁴	East (South)	291	27
1	1,000	9*10 ⁵	West	17,538	125
2	1,000	9*10 ⁵	East	7,030	71
1	4,000	3.6*10 ⁶	West	64,845	199
2	4,000	3.6*10 ⁶	East	83,415	128
	10,000	~10*10 ⁶	--	308,417	284

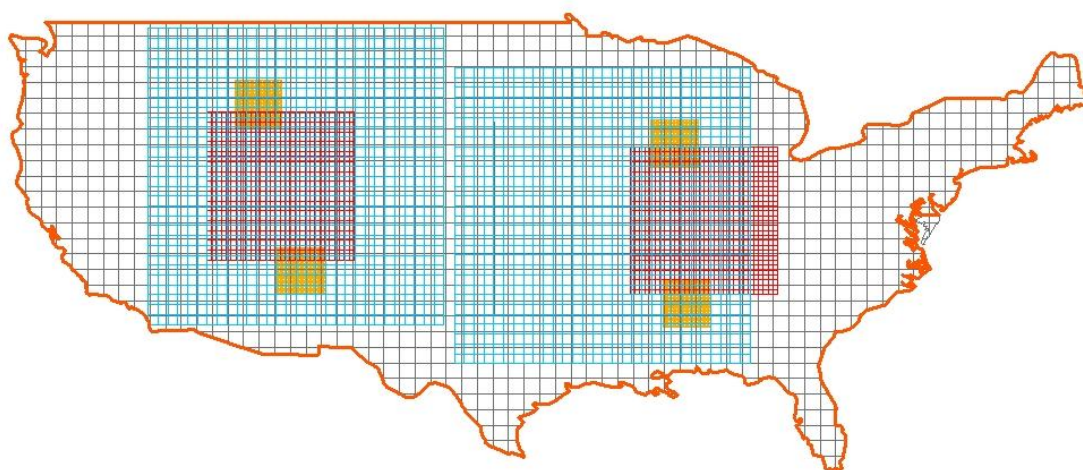


Figure 2.1: A map of the study area (the contiguous USA). Colored squares represent the sampling units. Orange sampling units have a grain of 100 km² and extent of 90,000 km²; red units have a grain size of 1000 km² and extent of 900,000 km²; blue units have a grain size of 4000 km² and extent of 3,600,000 km². Grain size of the grey grid extending the entire study area (~10,000,000 km²) is 10,000 km².

The major practical impediment for such analyses is data availability (Ferrier and Guisan 2006). Presence-absence data are only available for relatively small extents (Ferrier and Guisan 2006). Presence-only (occurrence) data have become available in the last decade at large quantities and for diverse taxonomic groups, via the internet (Graham et al. 2004). However, presence-only data are often considered improper for such analyses due to a range of inherent biases (Kadmon et al. 2004, Loiselle et al. 2008). The validity of using presence-only data in ecological analyses was studied several times in the context of modeling the distribution of a single species or modeling species richness patterns, but with inconclusive results (Elith et al. 2006). Chapter 1 evaluated the reliability of using presence-only data for studying multiscale diversity patterns based on the composition of taxonomic or functional groups. The assessment confirmed that presence-only data are sufficient for analyzing the relationships between species composition and environmental determinants. The objective of this chapter is to quantify the variation in the relationships between mammal species composition and its environmental determinants, at varying spatial scales.

Research hypotheses

At the base of the analyses performed in this chapter stand a series of hypotheses:

1. Climatic variables will explain the largest amount of variance in species composition, and their explanatory powers will increase with increasing spatial scale.
2. Within the climatic variables group, precipitation will account for the largest proportion of the explained variance.
3. Topography will have a negative correlation with scale, explaining more of the variance at the smaller scales of the analysis.
4. LULC variables will explain a substantial amount of variance at every spatial scale analyzed.

Methods

Data consisted of all occurrence records of terrestrial mammals (excluding bats) in the contiguous USA that existed in the GBIF portal (GBIF 2008). In addition to mammal occurrence data, Environmental data were compiled, including 15 parameters belonging to four groups (Table 2.2), which are believed to have an effect on species composition: climate; topography; land-use/land-cover (LULC) and primary productivity. The spatial resolution of all environmental layers was (or was reduced to) 0.0833° (~10km). Canonical Correspondence Analysis were applied to each of the sampling units and isolated the cumulative effect of each group of parameters with variation partitioning techniques (Cushman and McGarigal 2002, Legendre et al. 2005). Each parameter and each group was analyzed alone (marginal effect). In addition, groups were also analyzed with the other groups as co-variables (pure effect).

Results and Discussion

Climate had the largest effect on species composition at every spatial scale, and its effect increased with scale (Fig. 2.2). This is consistent with theoretical predictions that at fine scales, physical determinants' effects are obscured by biological interactions and that the effect of climate becomes more evident at larger scales (Wiens 1989 and references therein). LULC parameters decreased in their effect with increasing scale, while the effect of topography and NDVI remained relatively constant, and weak regardless of scale (Fig. 2.2). Analyzing the marginal effect of each parameter group resulted in higher values of explained variance, since it included the overlap in explained variance between parameter groups.

However, responses of the different groups to scale were similar, thus only the pure effects are presented hereafter. Differences in the amount of explained variance between sampling units within each scale were small relative to the difference between scales (Fig. 2.2). Such small inter-unit differences indicate that scale plays a major role in species-environment relationships, and that over

large geographic extents, it is more pronounced than regional or local differences in environmental conditions.

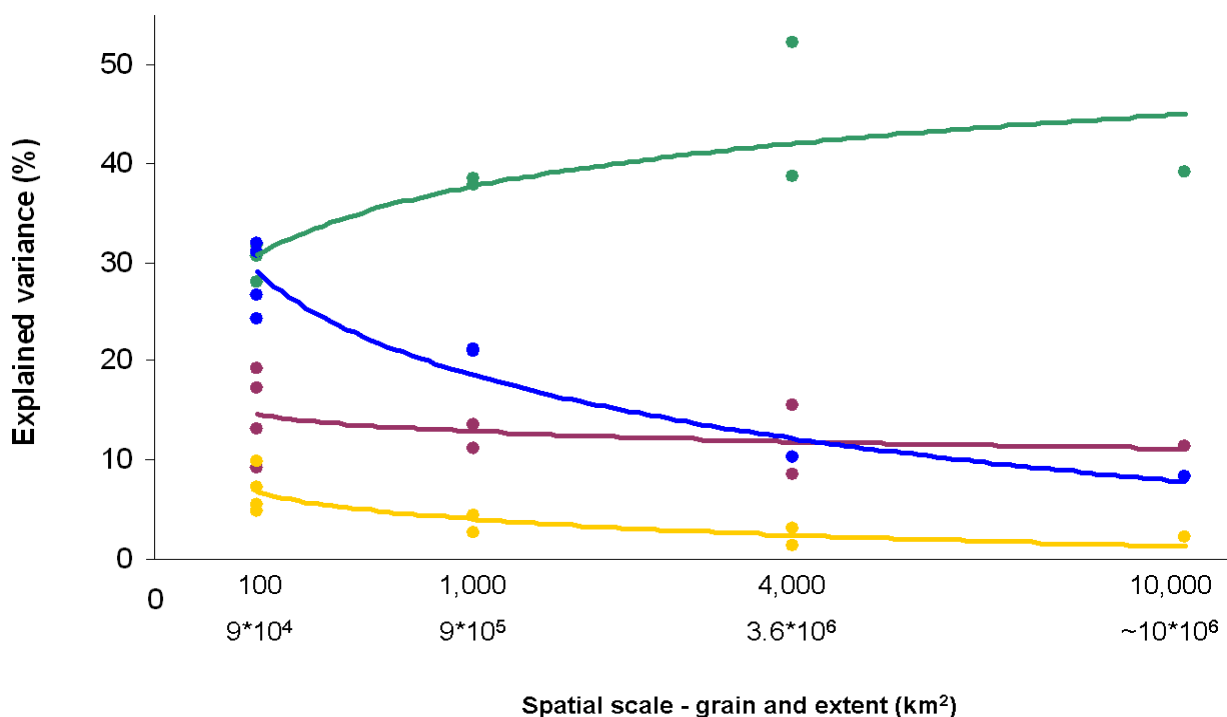


Figure 2.2: The amount of explained variance in the composition of mammal species in the contiguous USA, as resulted from CCA analyses, at varying spatial scales. Scale consists of grain size (upper number on the x-axis) and extent (lower number on the x-axis).

However, climate, topography and primary productivity at the third scale (4,000 and $3.6 \times 10^6 \text{ km}^2$ respectively), were higher in the western unit than the eastern unit (Fig 2.3), suggesting that there are significant differences between the Eastern and Western USA in major environmental gradients.

Wiens (1989) described a phenomenon called scale-domains, based on a review of studies that used different sized quadrates to study patterns of plant distributions. He suggested that change in pattern of ecological phenomena with scale is monotonous within each scale-domain. However, between domains pattern variability becomes chaotic and unpredictable, manifested as high variability

between sampling units. It is important to note that the effect of scale on the relationships between species and their environment is not directly by geographic extent, but rather by altering the length of the gradients which determine species distributions and composition.

Table 2.:2 Environmental parameters used in the analyses. * denotes binary)”dummy“(variables.

<i>Parameter group</i>	<i>Parameter name</i>	<i>Description</i>
Climate	Temperature	Mean annual temperature
	Temp_seasonality	Temperature standard deviation *100
	Precipitation	Mean annual precipitation
	Prec_seasonality	Coefficient of variation of precipitation
Topography	Altitude	Height above sea level
	Alt_range	Range of altitudes within grid-cell
Primary productivity	NDVI	Normalized difference vegetation index
Land-use Land-cover	Pop-density	Population density
	Urban*	An urban area
	Forestry*	Covered with forest
	Open-herbaceous*	Covered with herbaceous vegetation
	Agriculture*	Agricultural area
	Distance to urban	Distance to the nearest urban area
	Wetland*	Wetland area
	Water*	Covered with water

Breaking down the various parameter groups into individual parameters, I calculated the marginal effect of each parameter (Fig. 2.4). The general trend of the climate group followed that of mean annual temperature, and temperature seasonality (Fig. 2.4a), indicating that the dominant climatic feature is temperature, rather than precipitation, which explained a relatively small amount of data at the smallest scale, and an almost constant larger amount at the other scales. Precipitation seasonality had a small effect at every scale except the third. Pianka (1966) presented a theory of climatic stability, as a driving force of species diversity. The amount of variance explained by seasonality

measures in this study does not support that theory, since in most cases seasonality explained less variance than the mean value of both climatic features (Figure 2.4a).

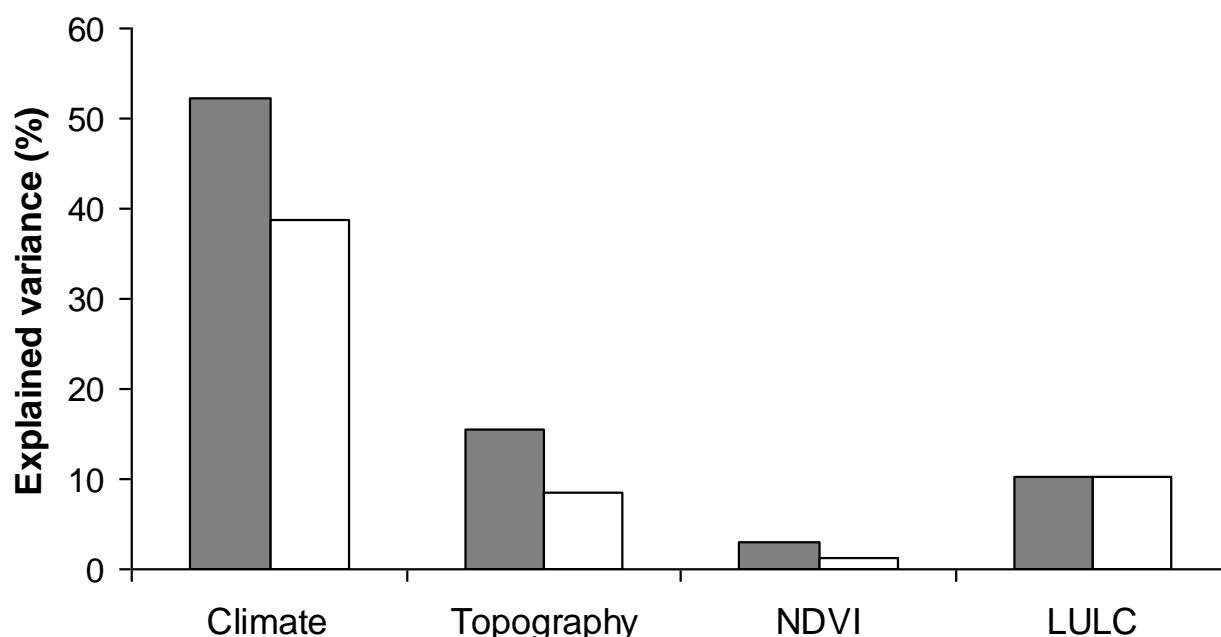


Figure 2.3: Amount of explained variance at a spatial scale of grain size and extent of 4,000 and $3.6 \times 10^6 \text{ km}^2$ respectively, by sampling unit. Grey bars are the values of the western unit; open bars are the values from the eastern unit.

In the topography group, altitude showed a constant increase in its effect, with increasing scale, while altitude range had a relatively small effect at all but the largest scale (Fig. 2.4b). At the largest scale, altitude range explained more variance than altitude itself. Such a switch in the importance of the topographic features may be of importance when selecting environmental parameters for modeling or reserve network planning. Among the theories of species diversity drivers presented by Pianka (1966) topographic relief, i.e. topographic variability, is suggested as a factor that increases species diversity. This study indicates that this theory might be valid at the largest scale. However, at smaller scales, at which topography plays an important role in determining species composition, altitude is more important than its range. All land-cover parameters in the LULC group were grouped together (Table 2.2). These parameters demonstrated a constant moderate decrease in their effect at the three smaller scales (Fig. 2.4c). At the largest scale, the amount of variance explained by all the grouped variables

dropped to ~3%. This is probably a result of the fact that at very large grain size, each grid-cell contains a mosaic of LULC patches, and the variability within the cells is equal to or larger than the variation between cells. Effect of population density decreased consistently with increasing scale, and distance to urban area followed a similar trend, except at the 100 km scale, where it explained a relatively large amount of the variance in species composition (Fig. 2.4c), suggesting that at the continental scale, the most important human factor is human presence, and not any specific human activity or behavior.

Analyses revealed that species composition is affected largely by climate at grain size between 10^2 to 10^7 km² and extent between $9 \cdot 10^5$ to 10^8 km² respectively. Topography was not a prominent factor in my analyses. It is probably more important at smaller scales (Pianka 1966). LULC parameters had sizeable influence on species composition at the two smaller scales, probably via habitat degradation and fragmentation, and ultimately, habitat loss (Wilson et al. 2004).

These results supply a quantitative indication that human induced global changes have an imminent impact on species composition, either directly or indirectly, and relate to the need for "...improved understanding of the response of biodiversity to changes in climatic factors and other pressures" as stated in a report by the Intergovernmental Panel on Climate Change (Gitay et al. 2002).

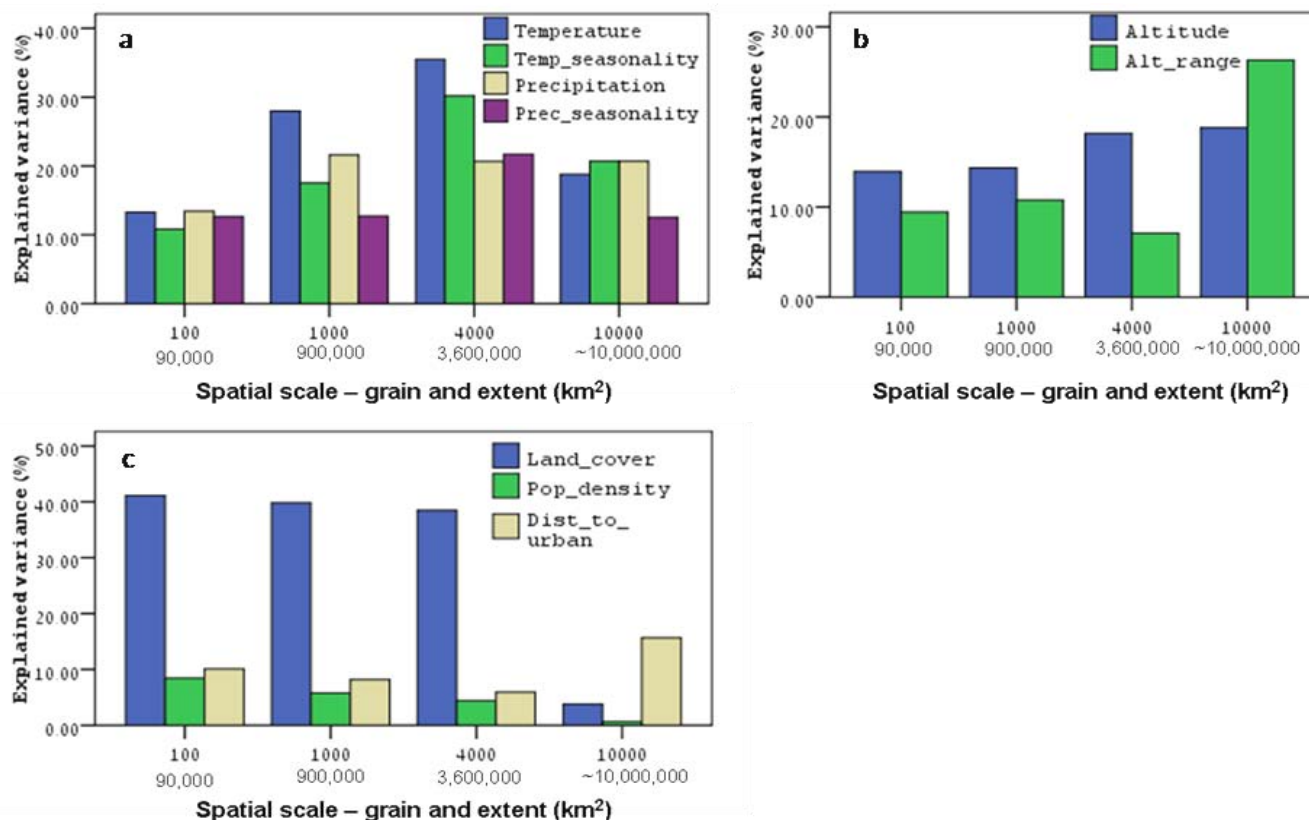


Figure 2.4: Amount of the variance in species composition explained by individual parameters: a) climatic parameters; b) topographic parameters and c) LULC parameters. Land cover is the combined effect of six land cover categories (agriculture, forestry, open herbaceous, urban, water and wetland). Prec_sea and Temp_sea stand for precipitation seasonality and temperature seasonality respectively; Dist_to_urb stands for distance to nearest urban area; and Pop density stands for population density.

As ecologists strive to find a general model explaining species diversity, the results presented here should be compared to other taxonomic groups and in other regions of the world. Nevertheless, this study, to the best of my knowledge, is the first to analyze the relationships between species composition and the environmental conditions that mold them at large and multiple spatial scales, contributing to the understanding of global patterns of biodiversity.

Chapter 3 - Comparison of classification methods for producing surrogates for biodiversity

INTRODUCTION

Conservation of biodiversity requires extensive knowledge of the distribution of a myriad of species. Such knowledge is scarce, and collecting all necessary data is often prohibitively costly. One widespread solution is to use surrogates for biodiversity (e.g. Belbin 1993, Faith and Walker 1996, Ferrier and Watson 1997, Kati et al. 2004a, Fleishman et al. 2005, Sarkar et al. 2006). The main assumption of the surrogacy approach is that the distributions of target species (species which are the target of conservation plans) is in spatial congruence with other, easily measured or collected, features. There are two types of biodiversity surrogates. In biological surrogates, the distribution of a taxonomic group is used to predict distribution patterns of other groups (Dobson et al. 1997, Virolainen et al. 2000, Reyers et al. 2001, Garson et al. 2002b, Kati et al. 2004b). Groups of species used as surrogates in the literature are numerous (Carmel and Stoller-Cavari 2006, Rodriguez et al. 2006). Dobson et al. (1997) found that conserving endangered plant species in the United States resulted in a representation of a large proportion of endangered species from all other taxa they examined, but that the presence of endangered birds and herptiles were more indicative of overall endangered biodiversity in the area. Kati et al. (2004b) examined the value of six taxonomic groups as biodiversity surrogates at a local scale, by comparing the spatial congruence of species richness patterns of each group to the other groups. They concluded that woody plant richness patterns were the best indicator of overall biodiversity patterns in their study area. The scale at which Kati et al. (2004) worked is specifically suitable for conservation (a natural reserve 430 km² in size) making their study highly important for conservation applications. Other groups tested as surrogates include mammals, mollusks, arthropod, beetles and butterflies etc. (Kerr et al. 2000, Rodriguez et al. 2006). Environmental surrogates are classifications of an area into land parcels with similar physical characteristics (Ferrier and Watson 1997, Reyers et al. 2001, Leathwick et al. 2003, Oliver et al.

2004, Bonn and Gaston 2005, Trakhtenbrot and Kadmon 2005, Rodrigues and Brooks 2007, Arponen et al. 2008). An ongoing debate concerns the effectiveness of surrogates in predicting species assemblages (Faith and Walker 1996, Rodrigues and Brooks 2007). Ferrier and Watson (1997) proposed two ways to quantitatively assess the effectiveness of surrogates: (1) based on the number of species represented by a set of sites selected for conservation using different surrogates, and (2) on the level of correlation between the spatial structure of a surrogate and of the taxonomic group of interest. Rodrigues and Brooks (2007) reviewed a large number of studies of surrogate efficiency. They found that in 59% of the studies reviewed, surrogacy was found positive, i.e. performed better than random selection of conservation areas in representing target species. They conclude that this indicates a weak yet positive value of surrogacy. They also found that cross-taxon surrogacy is stronger in fresh water environments than in the terrestrial and marine realms. They also concluded that taxonomically closer groups constitute more efficient surrogates than taxonomically very distance groups. They found that about half the studies on environmental surrogates had positive results, leading to the conclusion that in general, biological surrogates are more efficient in representing biodiversity than environmental surrogates. Carmel and Stoller-Cavari (2006) compared environmental and biological surrogates on Mt. Carmel, Israel. They concluded that environmental surrogates performed similarly well as did biological surrogates, and that woody plants were the most efficient biological surrogate, similar to Kati et al. (2004b)

A common approach for producing surrogates for biodiversity is classification of an area into domains of similar characteristics, using cluster analysis (Faith and Walker 1996, Trakhtenbrot and Kadmon 2005). Clustering can be conducted with various partitioning or agglomerating methods (Everitt 1993, Legendre and Legendre 1998), based on similarity or dissimilarity measures. Several arbitrary decisions are made during this process, regarding the similarity measures, number of classes

and the specific clustering algorithm to be used (Everitt 1993). These decisions may largely affect the resulting surrogate map.

Here, five clustering methods commonly used in conservation planning were evaluated, representing three different approaches to clustering, average based methods (AVERAGE, CENTROID and Ward's minimum variance), object based methods (i.e. furthest neighbor) and non-hierarchical classification (k-means). This study attempts to take steps towards understanding surrogacy in two directions. First, a quantitative evaluation of the efficiency of different clustering methods for surrogate production was carried out. Both biological and environmental surrogates were assessed, by applying different algorithms to the same dataset. The second element of this study adds a new aspect to the evaluation of surrogate performance, measuring the evenness of surrogate classes, in addition to their species representativeness. Surrogates are tools for planning reserve networks; classes covering very small areas are more difficult to incorporate into such networks. Surrogates with low evenness values are characterized by few dominant classes, and other classes occupying a negligible area (see for example fig. 3.1). In surrogates with a high evenness value, each class occupies a substantial part of the area. Such surrogate maps allow managers more flexibility in choosing areas for conservation and thus, more considerations may be taken into account, such as development, connectivity and land costs (Figure 3.1). Simpson's diversity index was used as a measure of evenness.

In this section there were competing hypotheses regarding the efficiency of the different classification methods. The null hypothesis was that the various classification methods will perform equally in both efficiency measures. Based on previous studies, the H1 hypothesis was that the Centroid method will be the most efficient method for producing surrogateH .tesh 2ypothesis was that Wa'rd s method will be the most efficient method for producing surrogates.

METHODS

Five different clustering algorithms were used to produce environmental and biological surrogate maps of the same area in Mt. Carmel, using three independent datasets: environmental variables; geophytes distribution; and woody species distribution. The surrogates' efficiency in representing species was evaluated, and in addition the evenness of the maps that were produced was calculated, so as to evaluate the level of flexibility given to reserve network planners by each method.

In addition, in order to examine the generality and robustness of our results, data from Trakhtenbrot and Kadmon (2006) was used to evaluate the performance of environmental surrogates produced using different clustering algorithms for the entire state of Israel.

Study area

Mt. Carmel in northern Israel has an area of ca~ 280km² and mean elevation of 220m (Fig. 3.2). The climate is eastern Mediterranean, with mean annual rainfall of 650mm*year⁻¹ and temperature averages ranging between 11°C in January and 24°C in August. The most common soil types are Terra Rossa and Rendzina. Vegetation is comprised of eastern Mediterranean scrubland consisting of structurally rich and diverse vegetation communities (Naveh and Dan 1973a, Le Honerou 1981, Naveh and Kutiel 1986). These landscapes, commonly referred to as vegetation mosaics, are highly heterogeneous at a broad range of spatial scales, ranging between grain size as small as a few meters to landscape level scales (Naveh 1975, Shoshany 2000, Bar Massada et al. 2008). The fine-grained mosaic is characterized by woody patches, herbaceous clearings, exposed rock and bare ground (Perevolotsky 2002). A grid of 500*500 m was superimposed on the area, dividing it into 1145 cells. All analyses were carried out at this spatial scale.

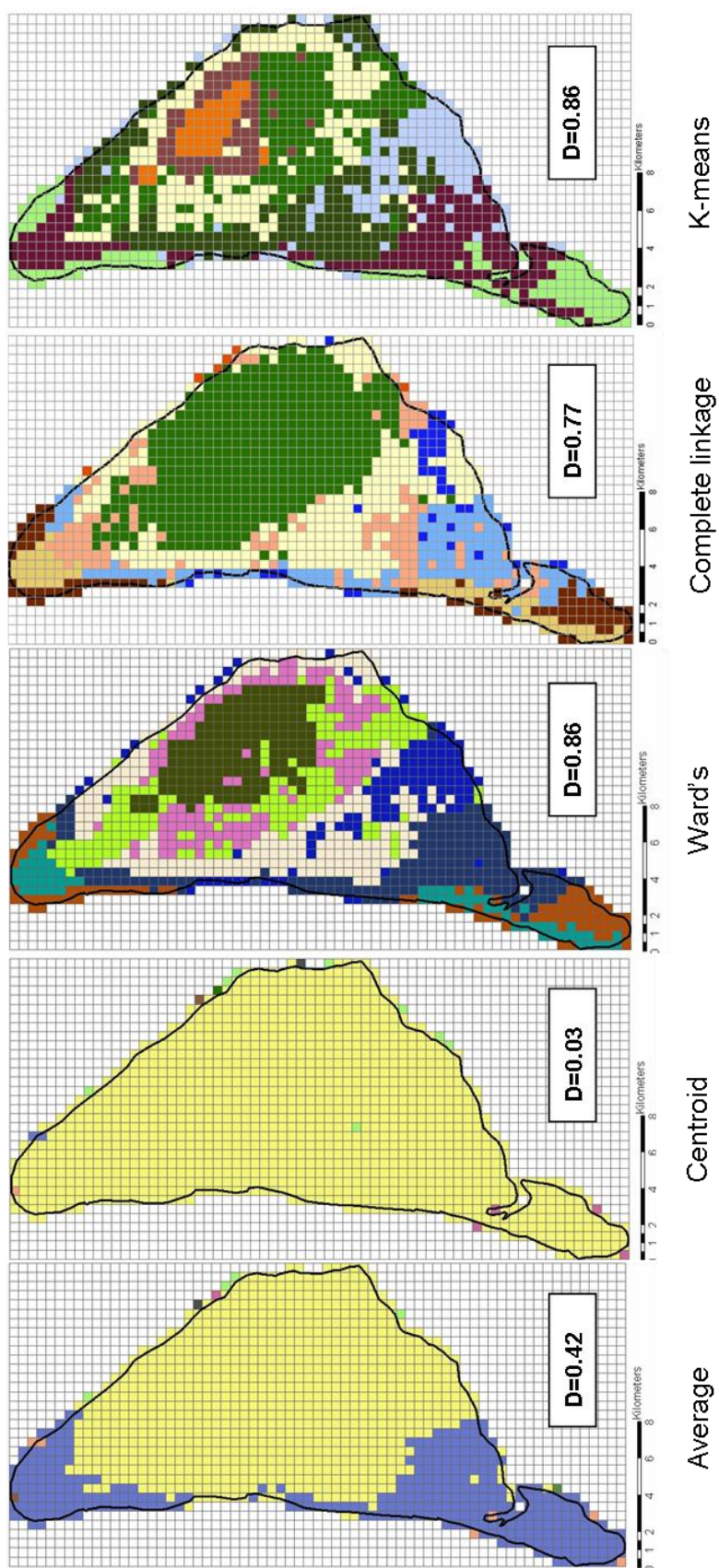


Figure 3.1: Biological surrogates with eight classes, based on woody species distribution, produced with five different clustering algorithms. D is Simpson's diversity index.

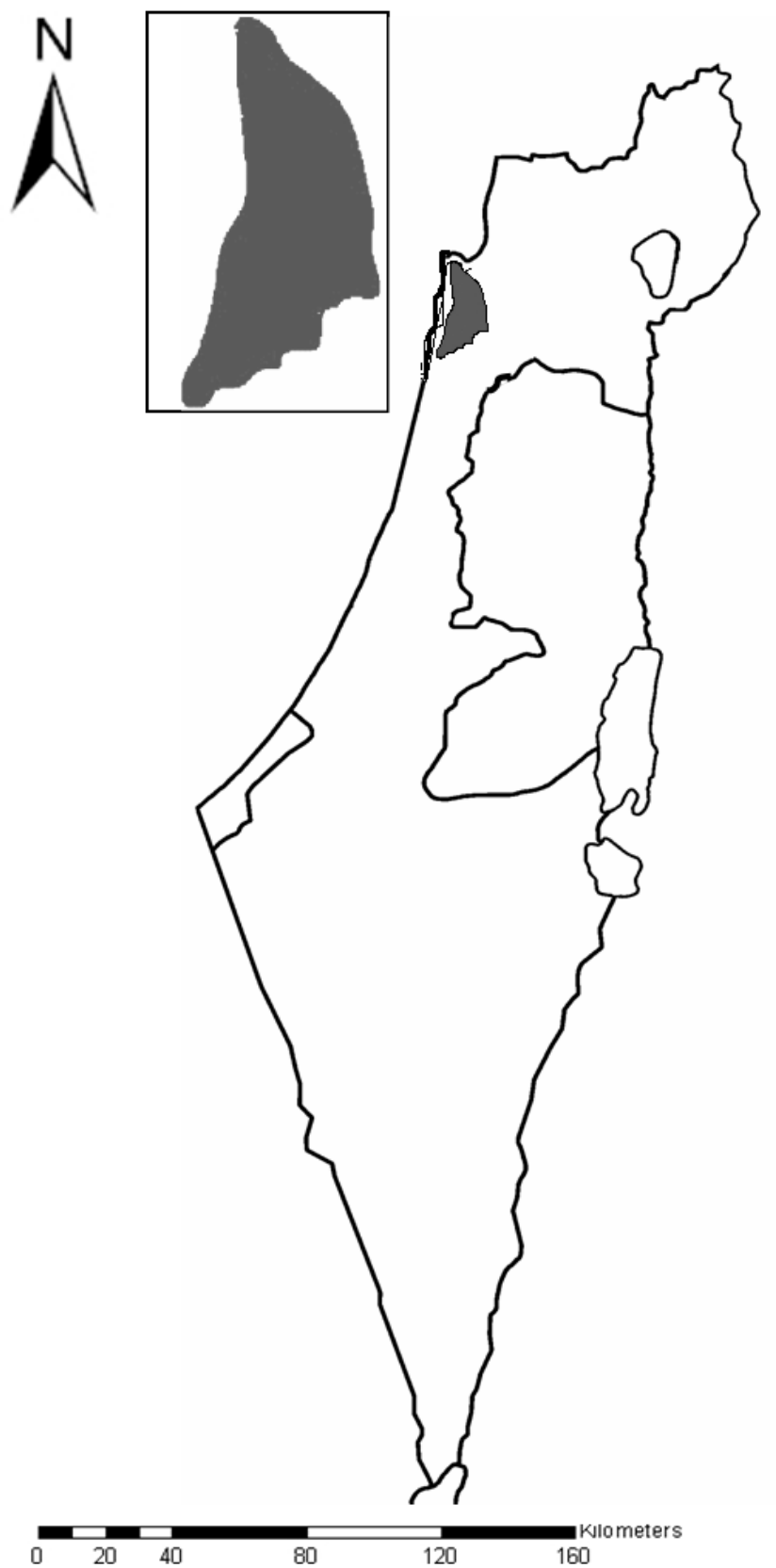


Figure 3.2: A map of the study area in northern Israel.

Species distribution data

Presence-absence data for geophytes and woody plants were collected in 100 sampling sites distributed randomly in the entire study area, from October 2002 through May 2003 (see Carmel and Stoller-Cavari 2006 for a complete description of fieldwork). Field data on each species, along with environmental parameters (Table 3.1), were used to create a habitat suitability map for each species. Habitat suitability maps were produced with logistic regression models (Guisan and Zimmermann 2000) which were applied to the study area in the geographic information system (GIS). Habitat suitability maps consisting of probabilities of occurrence, were transformed into binary (0/1) maps, with a threshold of 0.5. Statistically significant models were produced for 23 geophyte species and 37 woody species (Appendix 1). These sixty distribution maps were used in further analyses.

Table 3.1: Environmental parameters used for logistic regression models and environmental surrogates for biodiversity

Parameter	Description
NDVI	Normalized difference vegetation index – a measure of primary productivity
DTR	Distance to nearest road
Terra Rossa	Presence/absence of Terra Rossa soil in grid cell
Rendzina	Presence/absence of Rendzina soil in grid cell
Veg cover	Presence/absence of woody vegetation cover in grid cell
Rain	Average annual precipitation
mdt1	Mean daily temperature in the coldest month (January)
mdt8	Mean daily temperature in the hottest month (August)
Aspect	
Slope	Derived from digital elevation model
Elevation	

Environmental data

Environmental parameters used in this study included measures of climate, soils, vegetation, topography and an anthropogenic disturbance index (Table 3.1). All topographic parameters were calculated from a digital elevation model of the study area. Parameters were chosen to represent independent influences, i.e., climate has a different effect than topography, primary productivity affects species composition differently than distance to road or soil type etc. A total of 11 parameters were used, both for the logistic regression models (that produced species distribution maps), and for the environmental surrogates. All parameters were extracted from GIS layers and averaged over 500 m cells in order to fit the grid.

Surrogate production

Both types of surrogates, biological and environmental, were produced using five clustering algorithms: Average, Centroid and Ward's, complete linkage, and k-means. The first three methods are based on a calculation of some sort of average distance between groups of objects. In the Average method, the fusion of two object-groups is determined by calculating the distance between each pair of groups, as the average distance between each pair of objects within these groups. The two closest groups are fused together at each step, until there is only one group, containing all objects. In the Centroid method, the distance between group-pairs is determined by calculating the Centroid of each group, as an arithmetic average of the values of each of the environmental parameters. Ward's method is related to the Centroid method. It consists of calculating an error sum of squares (ESS), which consists of the average distance of each object in the group to the group centroid. The fusion of two objects is determined by finding the pair whose fusion contributes the least to the ESS (Everitt 1993, Legendre and Legendre 1998). The Complete linkage method is object-based, i.e., the fusion of two objects is determined by finding the two groups in which the distance between the two furthers objects is the smallest (Everitt 1993, Legendre and Legendre 1998). The four methods discussed

above are hierarchical agglomerative methods. The final clustering algorithm I used, k-means, is a non-hierarchical partitioning method. Similar to Ward's method, K-means partitioning is a least-squares method. However, it is a nonhierarchical classification method. It allows the user to divide a collection of objects into K groups (K is user-determined). Each parameter is represented as an axis in an n-dimensional space. The Euclidean distance between objects in this space is calculated, and then used to classify them into classes (Legendre and Legendre 1998).

Each clustering method was conducted using three classification schemes, with three, eight and twelve domains respectively. The two taxonomic groups (geophytes and woody plants) were used alternately, as surrogates for each other. In total, 45 different surrogates were examined.

Biological surrogates

Two similarity matrices were calculated, for woody and geophyte species using Jaccard's similarity coefficient. The similarity matrices were then used to classify the 1145 cells into domains. Three surrogate types were constructed: environmental surrogates, biological surrogates based on woody species distributions and biological surrogates based on geophyte species distributions. Each of these surrogate types was produced with three, eight and twelve domains, yielding nine combinations of surrogate type and number of domains. Each such combination was produced using all five different clustering methods: Average, centroid, Ward's, complete linkage and k-means. In total 45 surrogate maps were produced.

Environmental surrogates

Environmental surrogates were produced in the same manner as the biological surrogates, using GIS layers of environmental parameters instead of species distributions (Table 3.1). I chose Gower's similarity coefficient instead of Jaccard's in order to produce environmental surrogates, since it allows incorporation of both continuous and binary parameters, such as presence/absence of a soil type or

woody vegetation (Gower 1971, Legendre and Legendre 1998). In addition, when applied to binary data, Gower is equal to Jaccard, thus the two indices are compatible and interchangeable (Dunn and Everitt 1982).

Performance evaluation

Performance of the five different types of clustering was compared in two ways: (a) species representativeness - the number of species that are represented when a single site of each domain is chosen randomly and (b) Simpson's diversity index, in order to evaluate the evenness of the different domains within each surrogate map.

For species representativeness I recorded the cumulative number of species present in a set of sites, when a single site from each domain is chosen at random. Ten thousand random sets of sites were selected using a Monte-Carlo permutation procedure, one site from each domain. Environmental surrogates were evaluated using both taxonomic groups, while biological surrogates were evaluated against each other.

Simpson's diversity index was calculated for each set of surrogates, to test the diversity of the different classes as follows:

$$(1) \quad D = 1 - \sum_{i=1}^n \left(\frac{n_i}{N} \right)^2$$

Where n_i is the number of cells in class i and N is the total number of cells in the grid.

To evaluate the different clustering algorithms, The different algorithms were ranked between 1 and 5, for their performance compared to the other methods for each surrogate. Environmental surrogates were ranked according to their performance for each taxonomic group separately, and used the average rank for the final ranking. Biological surrogates were ranked according to their performance in representing species of the other group. Each surrogate type was ranked separately for each efficiency measure, and added them for total relative efficiency. In addition, nine environmental

surrogate maps of the entire flora of Israel, produced by Trakhtenbrot and Kadmon (2006) with three different algorithms (Average, Centroid and Ward's), and with 3 levels of partitioning (3, 8 and 12 classes) were evaluated. Trakhtenbrot and Kadmon report that the relative efficiency (representativeness) of the different algorithms was constant regardless of the number of classes. Average clustering performed best, followed by Centroid, and Ward's method was the least effective. Here, Simpson's index was calculated for these nine surrogate maps, and ranked the different maps accordingly.

RESULTS

Species representativeness

All types of surrogates, except one based on complete linkage and one based on k-means clustering, represented more species than a random choice of the same number of cells from the grid. Figure 3.3 shows the results of species representativeness for surrogates with eight classes. Surrogates with three and 12 classes showed a similar trend. The Centroid clustering method was the most efficient algorithm in representing species, representing 94% of the species on average, for the eight class scheme. This method was superior to the others in every combination of environmental surrogates, and in four of six combinations of biological surrogates. The Average clustering method was ranked highest in two combinations, representing an average of ~86% of the species, Ward's method was ranked third in species representativeness, representing an average of ~82% of the species for a scheme of eight classes. Complete linkage and k-means were the least effective methods for representing 79% and 81% of the species respectively. Both methods represented less species than a random choice of sites in one of the 9 surrogate combinations. Total ranks of the different surrogate types are summarized in table 3.3. Trakhtenbrot and Kadmon (2006) reported that, when applying a weighting scheme similar to the weights in our study, the method that ranked highest in species

representativeness of vascular plants in the entire state of Israel was Average, followed by Centroid, and that Ward's method was the least effective.

Table 3.3: Summary of the rankings of all different combinations of number of classes and clustering algorithms. Each algorithm was ranked 1-5 according to its relative performance in each number of classes, for each of the two performance evaluation measures (representativeness, Simpson's diversity index).

<i>Algorithm ranking</i>		
	SAI	Simpson
Average	32.5	18
Centroid	38.5	10
Ward	22	43
Complete linkage	23.5	25
K-means	19	39

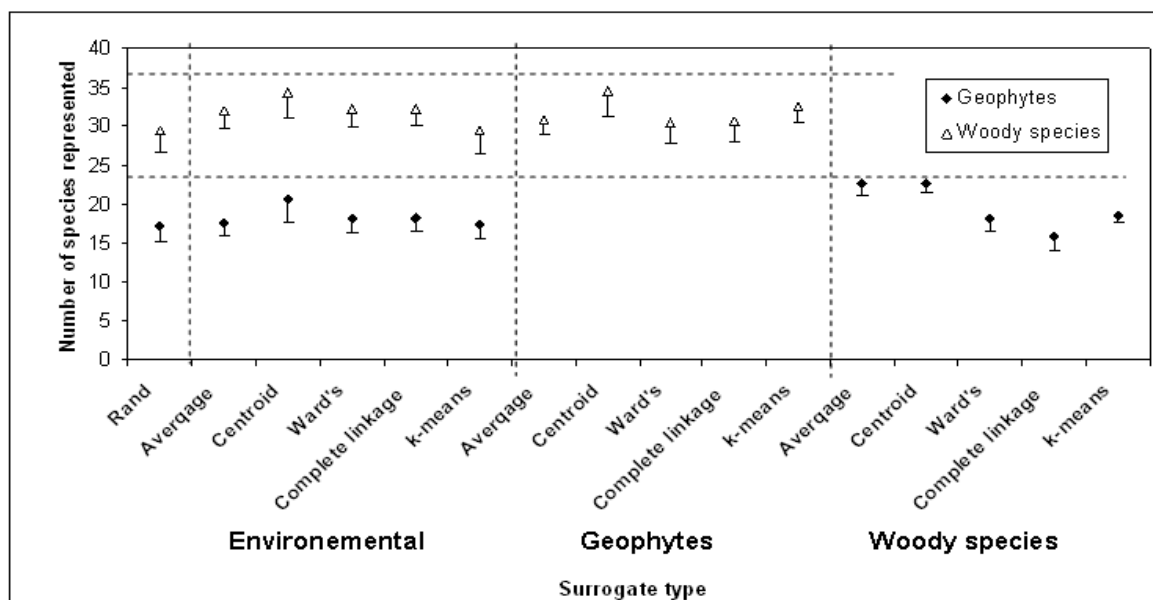


Figure 3.3: Average number of species represented by a set of surrogates for the eight classes scheme.

Surrogates based on geophytes were tested with woody plants only, and vice-versa. 'Rand' stands for a random selection of eight cells without classification, repeated 10 000 times. Values are average of 10,000 permutations of random selection of sites, one site per surrogate class. Error bars are standard deviations of the 10,000 permutations. Horizontal dashed lines represent total number of species of geophytes and woody species. Vertical dashed lines divide between different surrogate types.

Domains evenness

Figure 3.1 shows that the Centroid method effectively produced only one large class across the entire area (Fig 3.1), while the remaining seven classes occupied a very small area mostly near the perimeter. Four classes occupied only a single cell, and one class occupied two cells, with a value of Simpson's index of 0.03 (Fig 3.3), probably representing locations of rare species. Similarly, the Average method produced two large classes and six very small classes (Fig 3.1), with four classes occupying only a single grid cell, with a value of Simpson's index of 0.42 (Fig 3.4). Ward's method produced eight effective classes (Fig 3.1), with a value of Simpson's index of 0.86 (Fig 3.4). Ward's method produced surrogates with the highest Simpson's index value in seven out of nine cases (Fig 3.3).

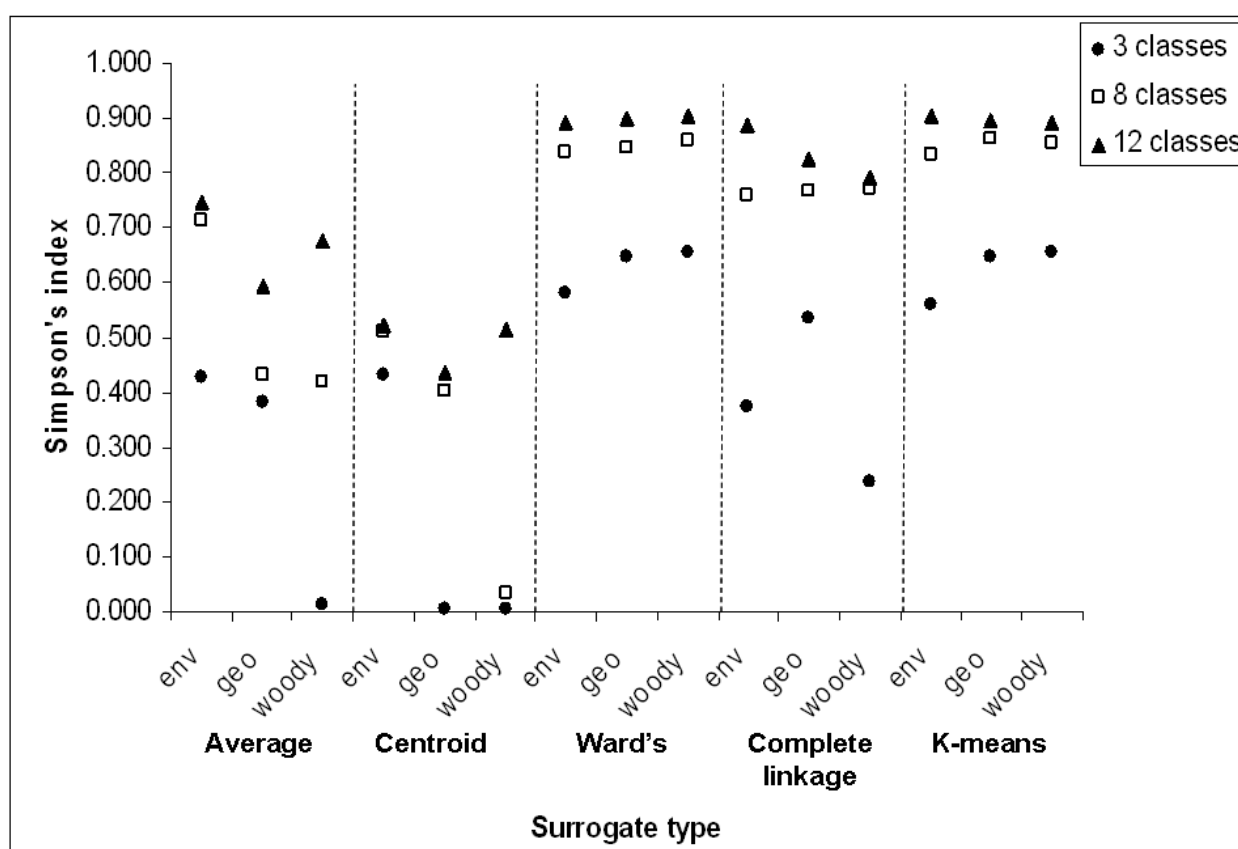


Figure 3.4: Simpson's diversity index for each type of surrogate.

Ranks given according to Simpson's index are summarized in table 3.3. Analyses of Trakhtenbrot and Kadmon's (2006) maps showed that Ward's method always produced maps with higher values of Simpson's index, and Average clustering was superior to Centroid in two of the three combinations.

DISCUSSION

Evaluating the efficiency of surrogates for biodiversity is a complex task that requires knowledge of the entire biological diversity (Rodrigues and Brooks 2007). A common solution is to use a group of known species as the target group, and to evaluate how well the surrogates represent them. Ferrier and Watson (1997) introduced a measure of efficiency, the species accumulation index, which measures how many species out of the entire pool are represented by a set of domains, and also compares it to a random selection of sites. Another measure they proposed was the Mantel correlation coefficient, which is the correspondence between domains and species distributions (Manly 2007). These measures do not fully account for an important aspect of surrogacy, namely, its actual use for site selection, and additional information is needed, such as the evenness of the different domains. If domains are relatively evenly sized, higher flexibility is allowed for site-selection between various options, whereas if some domains are represented by very few pixels (as in the case of the centroid algorithm output), little flexibility is left to managers in the site selection process. Evenness is an intrinsic trait, thus it is an independent measure of surrogate quality.

The various algorithms could be divided into two groups: the Average and Centroid methods, which represented more species than the other algorithms, but had low values of Simpson's index, and Ward's, complete linkage and k-means, which represented less species but had relatively high Simpson's values. Unfortunately, none of the algorithms evaluated excelled in both parameters. Current conservation practice takes place at two levels, coarse and fine filters (Margules and Stein 1989, Noss 1990, Pressey et al. 1993, Maddock and Du Plessis 1999, Schwartz 1999, Bonn and

Gaston 2005, Orme et al. 2005). Biodiversity surrogates are considered coarse filters, which are tools for capturing biodiversity in its broadest sense, including habitats, ecological processes and entire ecosystems, as well as individual species (Noss 1990). Fine filters concern rare or endangered species, which could have been missed by coarse filter methods. Surrogates with domains that cover very small geographic areas (low Simpson Index value), such as those produced by the Average and Centroid clustering algorithms, may be difficult for use as coarse-filters, due to the local nature of some of the resulting domains. However, since they indicate unique locations, particularly if the classification is based on species composition, such methods might be used to locate areas of high conservation value, and help to focus small scale conservation efforts.

The effect of the clustering algorithm on the efficiency of surrogacy has been largely overlooked, and very few studies quantitatively evaluated alternative algorithms. Tarkhtenbrot and Kadmon (2006) reported that the Centroid algorithm resulted in the best surrogates, which is similar to our results, when considering species representativeness alone. However, by calculating Simpson's diversity index in addition to species representativeness, I show that what seemed to be the better surrogate (centroid-based cluster analysis), may be a less effective tool for biodiversity conservation from certain perspectives, although the number of species it represents may be larger. In general, the results point out that there is no 'optimal solution' to surrogacy, and each case should be considered separately. In fact, results imply that there is a trade-off between species representativeness and domain evenness. This trade-off is shown to exist at more than one spatial scale, and for more than one set of environmental variables.

Knight et al. (2008) reported the gap between conservation science and conservation actions. These results indicate that a solution that may be considered scientifically superior may, in fact, be incompatible for managers. Figure 3.1 is an example of the difference between scientifically 'superior' and management-compatible solutions. In conclusion, scientific work on prioritization of areas for

conservation should include a final step, which is mostly disregarded in the literature, i.e. looking at the resulting maps, and examining them through a field conservationist's eye.

Chapter 4

Developing bio-environmental surrogates for biodiversity: A case study from Mt. Carmel, northern Israel

INTRODUCTION

Surrogates for biodiversity constitute a principal approach to conservation planning, as they are a means of using information on an easily measured (or calculated) feature as an indicator of the species assemblage in a given area (Ferrier and Watson 1997). Two types of surrogates exist.

Environmental surrogates are based on the partitioning of an area into domains with similar physical traits, usually by multivariate analyses, as proposed by Belbin (1993, 1995). Biological surrogates are based on the distributional data of a known group of indicator species. There is ongoing debate about the effectiveness of the different surrogates in representing true biodiversity patterns. For example, Araujo and Humphries (2001) claimed that there is no evidence for the ability of environmental surrogates to predict patterns of biodiversity, while Carmel and Stoller-Cavari (2006) conducted a direct comparison of the two types of surrogates, by measuring the congruence between patterns of the surrogates and of three groups of species. Environmental surrogates were produced with k-means classification of the area according to environmental conditions into environmental domains. They used analysis of similarity to test for surrogacy efficiency of each species group as surrogates for the other groups. They found that the most effective biological surrogates were woody species, which performed equally well as environmental surrogates. Dobson et al. (1997) concluded that endangered bird and herptile species are the most sensitive surrogates for other endangered species in the USA, based on spatial congruence between species distribution maps. Garson et al.

(2002a) tested the efficiency of breeding birds as surrogates of species at risk in Quebec. Their results indicated that bird species make relatively efficient surrogates in that system. In addition, they stated that the spatial scale of the analysis may affect its outcome. Kati et al (2004b) evaluated six groups according to their surrogacy efficiency, and found that woody species were the most effective surrogates, adequately representing all groups other than orchids. Lawler and White (2008) tried to characterize 'efficient' surrogates, according to a set of simple traits, i.e., taxonomic diversity; the nestedness of species distributions; biodiversity hotspots; species range sizes; and environmental diversity. They concluded that there are no simple characteristics that make effective surrogates. They did, however find that in general, biological surrogates performed better than environmental surrogates. In a review of 575 studies of both biological and environmental surrogates, Rodrigues and Brooks (2007) found generally positive, yet weak evidence that surrogates were effective, and that cross-taxon surrogates outperformed surrogates based on environmental data. They also relate to the source of the data used. They concluded that the best combination for testing surrogacy efficiency is using extrapolated data for the surrogates, and field data for the target species, since it minimizes commission and omission errors in the surrogates and targets, respectively. A method for the selection of reserve networks that is based on either environmental or biological data was presented by Faith and Walker (Faith and Walker 1996). They propose to calculate the 'environmental diversity' represented by a set of sites. Environmental diversity is calculated from an ordination space, allowing the use of actual environmental data, or inferred environmental data, derived from the ordination of distributional data. That approach offers an additional advantage - unlike clustering methods resulting in discrete units, the space from which sites are selected is continuous. It also allows calculating which of the remaining sites will contribute the most to the representativeness of the environmental diversity. A different approach combines the two types of surrogates to create a hybrid surrogate (Ferrier 2002, Ferrier et al. 2002, Sarkar et al. 2006). In their review of hybridization approaches,

Ferrier and Guisan (2006) related to the excess information existing in spatial representations based on both environmental parameters and best available biological data. They presented three approaches of community-level modeling. The first approach - 'assemble first, predict later' involves assembling of available survey data into 'community entities', and then modeling the distribution of these entities in relation to environmental predictors; 'predict first, assemble later', in which the first step is producing distribution models for all available species, stacking them, and aggregating them by means of classification or ordination; and 'assemble and predict together', in which all species are subjected to a single simultaneous modeling process in relation to environmental conditions. Here a new approach to the hybridization of biological and environmental surrogates, bio-environmental surrogates (BES), is presented. It includes the 'predict first, assemble later' approach has a first step of the analysis. This step involves creating a predicted habitat-suitability map for each surrogate species, and then dividing the area to domains of similar species composition. The next step was to directly combine the results of this step with environmental data in order to produce fully hybridized surrogates.

Ecological theory supports the concept of BES as possibly superior to both its precursor surrogates. Species occupy a realized ecological niche, a multi-dimensional space confounded by inter-specific interactions such as competition, predation and parasitism, as well as dispersal limitations and chance effects (Connell 1961, Vandermeer 1972, Wilson et al. 2003). I suggest that BES follow the same principal, constraining environmental domains according to biological patterns. The goal of this study is to develop a new approach for producing hybrid surrogates for biodiversity, based on both environmental and biological data and evaluate their efficiency in representing species, under the hypothesis that such hybrid surrogates will represent biodiversity more efficiently than single source surrogates, while maintaining relatively high levels of evenness, making them efficient tools for conservation practitioners.

METHODS

Study area

The study took place on Mt. Carmel, northern Israel, in an area of approximately 280 km². Climate in the study area is eastern Mediterranean, with mean annual rainfall of ~650 mm and average daily temperatures of 11°C and 24°C in January and August, respectively. Vegetation is defined as eastern Mediterranean scrubland, consisting of structurally rich and diverse vegetation communities (Naveh and Dan 1973b, Naveh and Kutiel 1986). These vegetation mosaics are highly heterogeneous at a broad range of spatial scales, ranging from grain size as small as a few meters to landscape level scales (Naveh 1975, Bar Massada et al. 2008). The mosaic is characterized by woody patches, herbaceous clearings, exposed rock, and bare ground (Perevolotsky 2002). The study area was divided into 1145 cells by a superimposed grid with mesh size of 500 x 500 m.

Species distribution data

Presence-absence data for geophytes and woody species were collected in 100 sampling sites distributed randomly across the study area from October 2002 to May 2003 (see Carmel and Stoller-Cavari 2006 for a complete description of fieldwork). These groups were used in the analyses for two main reasons. First, there was high-quality presence –absence data with good spatial coverage of the study area for these groups (Carmel and Stoller-Cavari 2006). Second, vascular plants in general, and woody species in particular were used successfully as surrogates for biodiversity in the past (Chiarucci et al. 2000, Pharo et al. 2000, Kati et al. 2004b). These data were used, along with several environmental parameters (Table 4.1) to produce habitat suitability maps for 23 geophyte species and 37 woody species (Appendix 1). Habitat suitability maps, consisting of probability of occurrence for each species, were produced by applying logistic regression models (MacNally et al. 2003) to the study area. Species distribution maps were produced by assigning presence/absence values to each

cell, according to its probability, with a threshold of 0.5. Model parameters were selected using forward stepwise selection, with a significance level of <0.05 . Ferrier et al. (2002) referred to this approach as 'single first-classification later' community level modeling.

Table 4.1: Environmental parameters used in the study for logistic regression models and environmental surrogates for biodiversity

<i>Parameter</i>	<i>Description</i>
NDVI	Normalized difference vegetation index – a measure of primary productivity
DTR	Distance to nearest road
Terrarossa	Presence-absence of Terrarossa soil in grid cell
Rendzina	Presence-absence of Rendzina soil in grid cell
Veg cover	Presence-absence of woody vegetation cover in grid cell
Rain	Average annual precipitation
mdt1	Mean daily temperature in the coldest month (January)
mdt8	Mean daily temperature in the hottest month (August)
Aspect	Topographic aspect derived from a DEM
Slope	Topographic slope derived from a DEM
Elevation	Elevation derived from a DEM

Environmental data

Environmental parameters used in this study included parameters with expected ecological significance, such as measures of climate, soil type included two dummy variables of the two most common soils in the study area, i.e. Rendzina and Terra-rossa, vegetation, topography, and distance to the nearest road as a measure of anthropogenic disturbance (Table 4.1). Topographic parameters were derived from a digital elevation model of the study area. A total of 11 parameters were used (Table 4.1), both for the logistic regression models mentioned above, and as the basis of the surrogates for biodiversity (environmental and BES). All parameters were extracted from Geographic Information System (GIS) layers and averaged over each of the 500 m cells.

Production of surrogates for biodiversity

All types of surrogates and precursors were produced using hierarchical cluster analysis with Ward's method of minimum variance (Ward 1963, Legendre and Legendre 1998), to partition the study area into domains with similar species composition (biological surrogates), or similar environmental conditions (environmental surrogates). Ward's method produces domains that are more evenly distributed than other clustering methods, rendering it more useful for management purposes, as demonstrated in chapter 3. Cluster analyses were carried out with Matlab 7.3 (MathWorks, Natick, Massachusetts, USA).

Biological surrogates

In order to produce biological surrogates from the presence-absence data, a similarity matrix was calculated using the Jaccard similarity index for each pair of cells in the study area. Each functional group (geophytes and woody species) was used to produce a separate set of domains, which was then evaluated against the other functional group.

Environmental surrogates

The Gower similarity index was used as the basis for the environmental cluster analysis (Trakhtenbrot and Kadmon 2005, 2006). The Gower similarity index is most suitable for describing relationships between cells in a multidimensional space created by numerous environmental parameters. It is a robust metric, which performs well on both continuous and categorical parameters (Gower 1966, 1971, Legendre and Legendre 1998). Dunn and Everitt (1982) found that Gower similarity index was equal to Jaccard similarity index when binary data was used. Environmental surrogates were produced through the same process as biological surrogates. Performance of environmental surrogates was evaluated against both functional groups (geophytes and woody species).

Bio-environmental surrogates (BES)

Bio-environmental Surrogates (BES) are hybrid surrogates for biodiversity constructed using both environmental and biological data. Hybrid surrogates have been proposed in the past (Ferrier 2002, Ferrier et al. 2002, Sarkar et al. 2006) as possibly superior to either type alone.

Table 4.2: The different types of surrogates for biodiversity

	Surrogate name	Description
Single source	Environmental	Environmental surrogate, based on cluster analysis using Ward's method of minimum variance
	Geophytes	Biological surrogate, based on cluster analysis of geophytes' distribution using Ward's method
	Woody species	Biological surrogate, based on cluster analysis of woody species distribution using Ward's method
Hybrid	Niche_geophytes	Niche BES using geophytes distribution
	Niche_woody_species	Niche BES using woody species distribution
	Reverse niche_geophytes	Reverse niche BES using geophytes distribution
	Reverse niche_woody_species	Reverse niche BES using woody species distribution
	Intersection_geophytes	Intersecting BES using geophytes distribution
	Intersection_woody_species	Intersecting BES using woody species distribution

Here three alternative approaches for combining biological and environmental information to produce BES are proposed. The first approach hereafter the 'niche approach' is a hierarchical two-step classification of the area into domains. The first step is to produce environmental domains, each of which is then divided into sub-domains using biological information. The biological sub-domains are produced for each environmental domain independently by cluster analysis of the species present within each domain. Three environmental domains were produced, and three biological sub-domains for each environmental domain, yielding a total of nine domains. The 'reverse-niche approach' employs the same classification routine as described for the 'niche approach', above, only in reverse order. The study area is first classified into biological domains, and then each domain is subdivided using environmental information. In the 'intersection approach', the first step is to create environmental and biological domain layers, independently. The second step is an intersection of the

two layers to form the hybrid domains. All surrogate maps, including the biological, environmental and hybrid surrogates, comprised nine domains.

Performance evaluation

Two parameters were used to evaluate surrogate performance, representativeness and evenness.

Species representativeness was evaluated as the average number of species of geophytes and woody species represented by a random selection of a single site from each domain, which was calculated for each surrogate type, using a Monte-Carlo permutation procedure with 10,000 repetitions using Matlab. Performance measures of the three different hybridization approaches were compared to each other and to environmental and biological surrogates. The results were compared to a random selection of nine sites from the entire study area using the same Monte-Carlo procedure, in order to examine whether the produced surrogates were better than random. In order to evaluate the level of evenness of the domains within each surrogate map, Simpson's diversity index was calculated for each surrogate map.

The biological surrogates and BES were produced once using geophytes data, and once using woody species data. Surrogates based on geophytes were evaluated against woody species, and vice versa. Environmental surrogates were evaluated against both functional groups.

Surrogate fragmentation

Fragmentation metrics were calculated at the landscape level, i.e. the total number of patches in the surrogates maps, and the level of cohesion of the entire map, using Fragstat 3.4 (McGarigal et al. 2002). Cohesion is calculated as follows:

$$COHESION = \left[1 - \frac{\sum_{i=1}^m \sum_{j=1}^n p_{ij}}{\sum_{i=1}^m \sum_{j=1}^n p_{ij} \sqrt{\alpha_{ij}}} \right] \left[1 - \frac{1}{\sqrt{A}} \right]^{-1} * (100)$$

Where p_{ij} = perimeter of patch ij in terms of number of cell surfaces; α_{ij} = area of patch ij in terms of number of cells; and A = total number of cells in the landscape. It is a dimensionless measure of the overall connectedness of the various components of the map. High cohesion levels indicate low fragmentation.

RESULTS

Surrogate representativeness

A random selection of nine sites resulted in an average representation of 17.6 of 23 geophyte species and 30.1 of 37 woody species (Fig. 4.1). When a single cell of each domain was selected, biological surrogates based on woody species represented an average of 18.6 geophyte species. Surrogates based on geophytes represented 31.8 woody species (Fig. 4.1). Environmental surrogates performed similarly to biological surrogates. Selection of a single cell from each domain of environmental surrogates, produced in the same manner as the biological surrogates, represented an average of 18.4 geophytes and 32.3 woody species (Fig. 4.1).

BES were produced using three different approaches, and all three surrogates represented considerably higher numbers of geophytes and woody species than both biological and environmental surrogates. The Niche BES domains represented 36.98 woody species, while both reverse-niche and intersection BES domains represented 36.8 of 37 woody species. All three BES approaches resulted in representation of 22.9 of 23 geophytes species (Fig. 4.1). The differences are significant for both geophytes (one-way ANOVA, $F=38,000.062$, $p<0.001$) and woody species (one-way ANOVA, $F=26,449.102$, $p<0.001$).

Surrogates fragmentation

A qualitative visual examination of the resulting maps (Fig. 4.2) revealed that BES were characterized by relatively uniform and continuous domains, while domains produced using single-

source surrogates were more fragmented and irregularly shaped. Simpson's diversity index values ranged between 0.81 for intersection-BES based on geophytes to 0.89 for biological surrogates based on woody species. Due to the high values and small differences, all surrogate types were considered as equal in the evenness evaluation. All biological surrogates had more patches and lower cohesion values than the environmental surrogates.

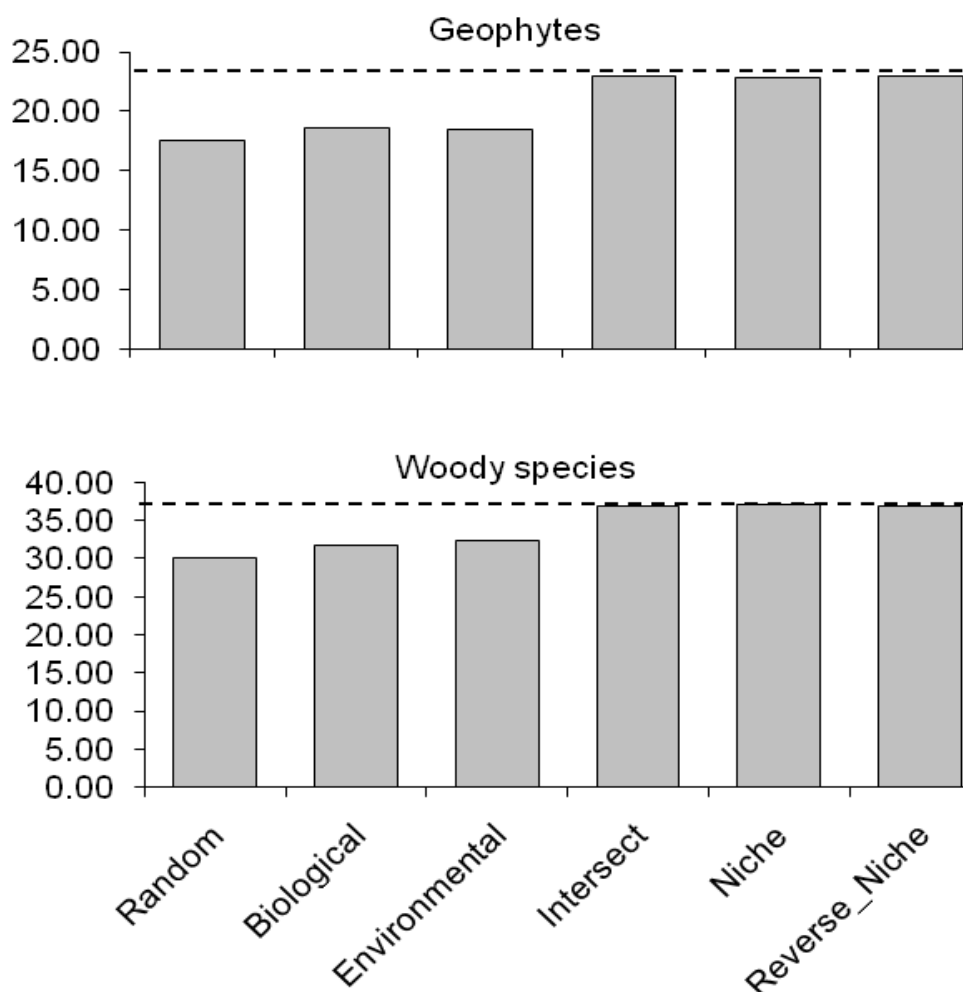


Figure 4.1: The average number of geophytes (top) and woody species (bottom) represented by each surrogate, calculated by applying a Monte-Carlo randomization procedure with 10,000 permutations.

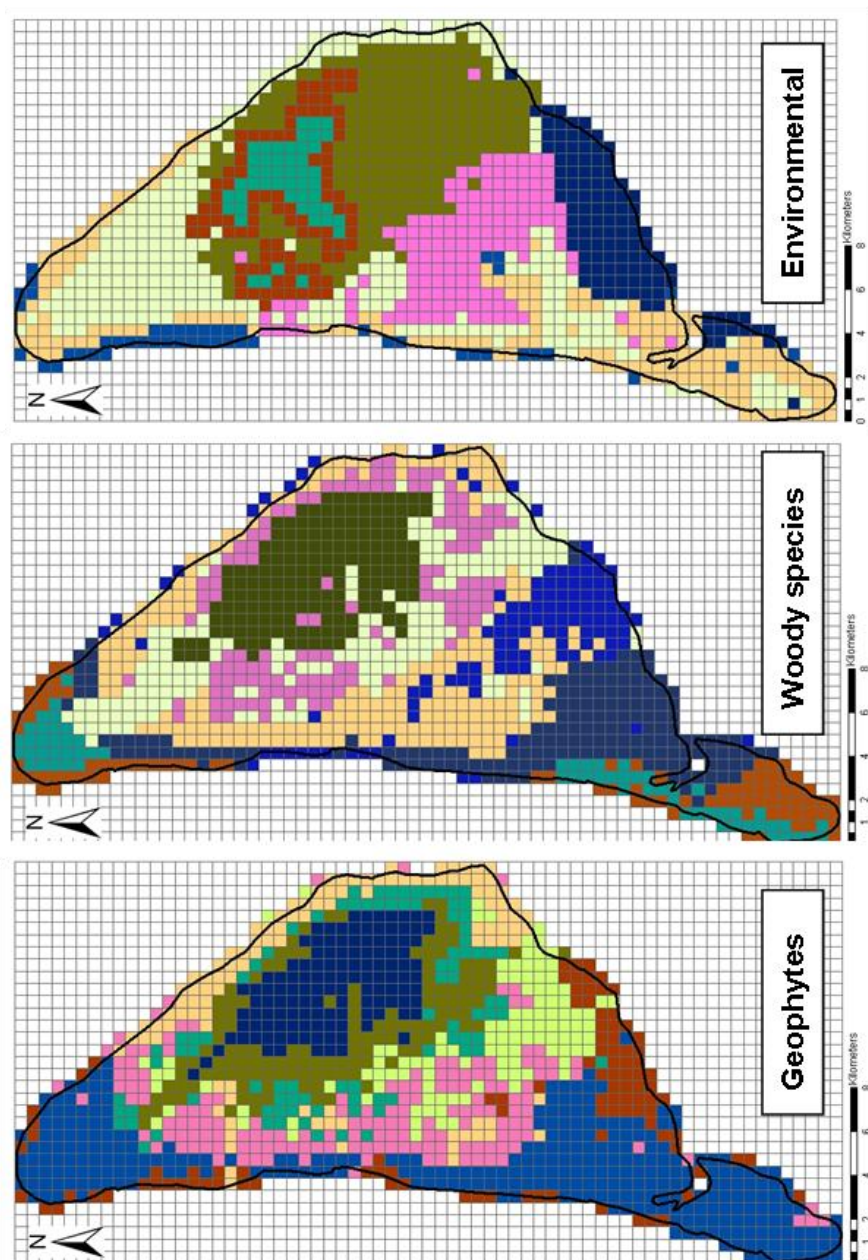


Figure 4.2: Comparative maps of surrogates for biodiversity on Mt. Carmel, northern Israel. The top three maps are single-source surrogates (either environmental or biological); the other six are hybrid surrogates. Each surrogate map consists of nine units, represented by the different colors. Surrogate name codes are specified in Table 4.2.

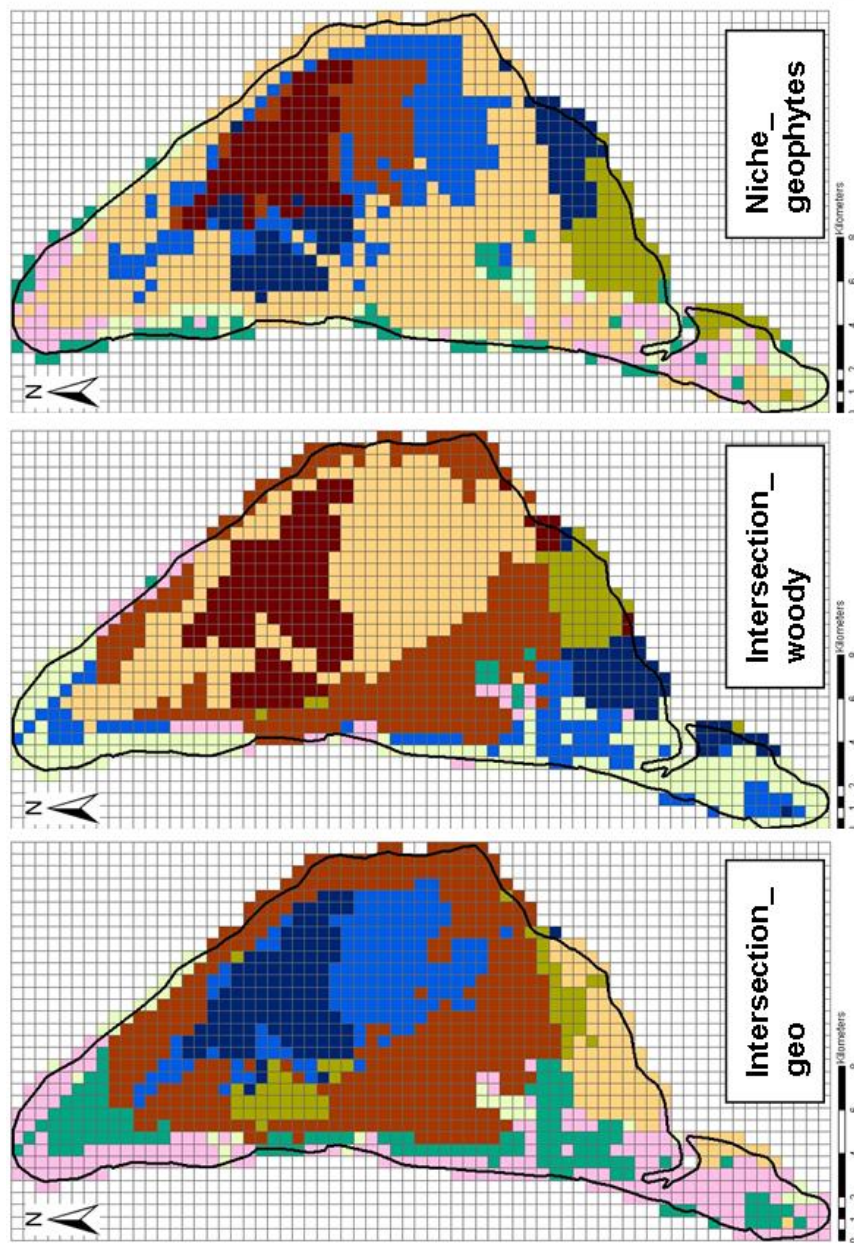


Figure 4.2 continued

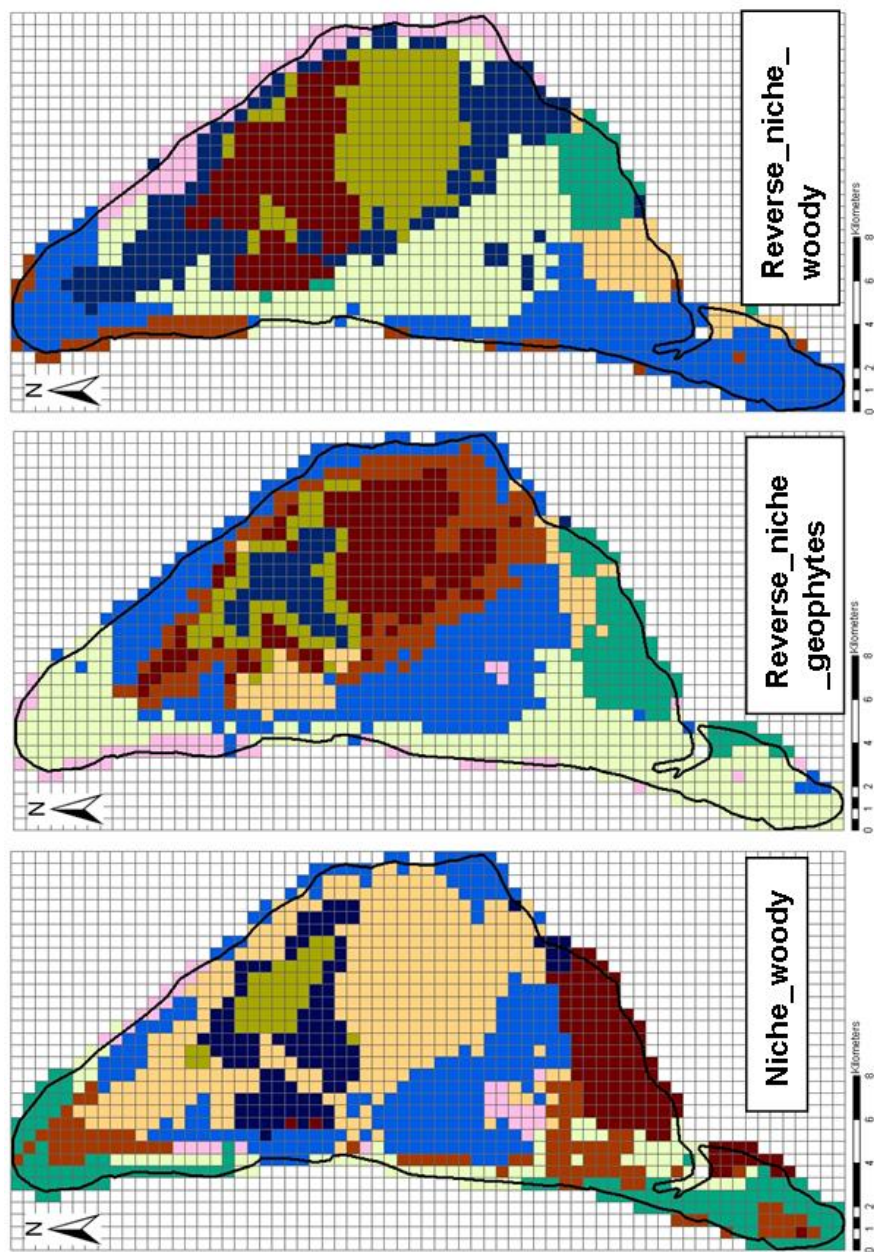


Figure 4.2 continued

BES values of both the number of patches and cohesion were similar to those of the environmental surrogates (Fig. 4.3). The only exception was niche-BES based on geophytes, which had a cohesion level similar to the other BES, but the number of patches was between BES and the biological surrogates (Fig. 4.3).

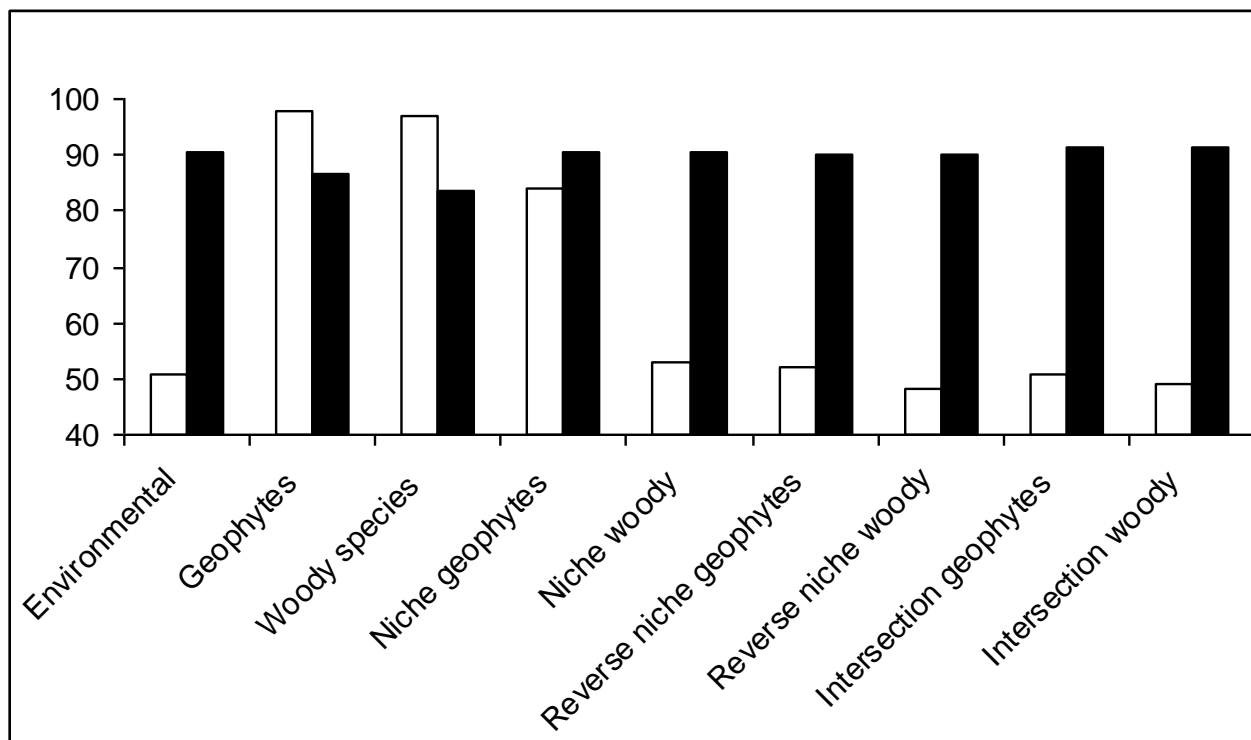


Figure 4.3: Measures of fragmentation of the different surrogate types. Open bars are the number of patches in the maps. Black bars are the cohesion level of the maps.

DISCUSSION

The results confirmed that hybridization of surrogates for biodiversity, i.e. the use of environmental and biological information for classifying the area into domains, increased the number of species that can be represented in a reserve network of a certain size, compared to single-source surrogates.

Although this concept has been suggested before (Kirkpatrick and Brown 1994, Ferrier and Watson 1997) but was never applied in practice. This work presents a novel approach to surrogate hybridization, with quantitative indication of its superiority over single-source surrogates.

The principal advantages of BES are best understood in light of the ecological niche theory (Hutchinson 1957). Environmental surrogates resemble a potential ecological niche of a species assemblage. Biological surrogates contain information on the confounding effects of the niche, but they lack explicit information on its physical traits. Hybridization of these two information sources

may combine both information types into an equivalent of the realized niche of the species assemblage (Ferrier et al. 2009). The concept of hybrid surrogates for biodiversity has existed for at least a decade (Ferrier and Watson 1997). The main obstacle along the way to producing them has been the differences between continuous environmental data and binary species distribution data. By applying the same clustering algorithm, i.e. Ward's method (Ward 1963, Legendre and Legendre 1998), to both datasets and using interchangeable measures of similarity (Gower and Jaccard coefficients for environmental and biological data, respectively), compatible spatial domains were produced. BES contains information on both environmental filters and the biological factors refining them.

Environmental parameters were also used to produce environmental surrogates for modeling species distributions for the biological surrogates. These affected the hybrid surrogates twice, indicating possible redundancy. However, the impact of environmental parameters on the spatial distribution of biological surrogates was different from the respective impact on environmental surrogates, since different sets of environmental parameters were selected for each species model, and environmental parameters affected spatial patterns of different species in different manners. In any case, such redundancy, if existed, would have acted to reduce the predictive power of BES. Thus, the conclusions regarding the superiority of BES over biological and environmental surrogates are not likely to change due to this potential redundancy.

In conclusion, planning reserve networks with the aim of conserving biodiversity effectively is complex, often requiring both financial and land resources, which are very difficult to come by. Thus, in recent years, conservation biologists have increased their efforts to develop more effective surrogates for biodiversity, in order to mitigate the need for resources, and to shorten the time needed to collect sufficient data to make conservation decisions (Margules and Stein 1989, Maddock and Du

Plessis 1999, Schwartz 1999). Hybridization of environmental and biological surrogates seems to be a step towards making surrogacy more accurate, and thus more efficient and robust.

CHAPTER 5: SYNTHESIS

This work concerns scale-dependent species – environment relationships, and how they can be incorporated into conservation science in order to increase the efficiency of the limited resources available for biodiversity conservation. As the relationships between species composition patterns and environmental variables are scale dependent (Wiens 1989, Levin 2000) environmental surrogates, which are based on the concept of the ecological niche and represent the various niches occupied by the species in the area of interest, are also scale dependent. Thus, it is problematic to assume that surrogates derived from remotely sensed data at large spatial scales will be efficient in predicting local scale patterns in biodiversity distribution. I propose that analyses aimed at producing surrogates for biodiversity should be conducted at a scale appropriate for the conservation of the target species. Such scale-specific surrogates may be produced by quantifying the effect of the various environmental variables affecting the assembly of species in the area of interest. Using the results from analyses such as those presented in chapter 2 of this work in order to create a weighting scheme of environmental variables that enter the analyses of environmental or hybrid surrogates may result in accurate scale-specific surrogates more efficient than surrogates produced using non-weighted or arbitrarily weighted environmental variables.

The different chapters of this work relate to the research objectives presented in the introduction section which address two central questions related to species-environment relationships. The first question relates to the understanding of the relationships between species and their environment, and how they are affected by spatial scale. Spatial scale is important here not only due to geographical distances involved, but rather a change of scale alters the length and shape of the environmental gradients involved in species composition assembly, as well as the species that are affected by them

(Ferrier and Watson 1997). From the results of the analyses in chapter 2, it is clear that environmental parameters have differential responses to changes in spatial scale. Theory suggests that climate is the most important environmental variable at large spatial scales, and that its effect increases with spatial scale, since climate is heterogeneous at large scales, while at smaller scales, it is relatively homogeneous (Menge and Olson 1990, Palmer and Dixon 1990). Results presented in chapter 2 give a quantitative affirmation to that theoretical prediction. In addition, breaking down climate into individual variables (temperature, precipitation and the seasonality of them) revealed that temperature is the single variable accounting for the largest proportion of explained variance in mammal species composition in the USA. The answers to the main research questions, as presented in this work, may help conservationists in their attempts to protect biodiversity, as demonstrated by the example given below. The second major research question derived from the specified research objectives relates to the ability to increase the efficiency of surrogates for biodiversity in conservation planning, by relating species-environment relationships from one group of species to another, and inferring on its spatial distribution. Results presented in chapters 3 and 4 indicated that by using certain classification methods, and by combining biological and environmental information, the efficiency of surrogacy may increase considerably. Since surrogacy is based on the ecological niche concept (Margules et al. 2002), the quantification of the species-environment relationships can be directly related and incorporated into the surrogate production process, rendering it more efficient and more accurate than existing surrogacy approaches.

An example of that concept is given - Consider for example the selection of areas for a reserve network for the conservation of mammalian diversity in the contiguous USA, and consider that the grid has a cell size of 100 km^2 . Using either environmental surrogates (chapter 3) or BES (chapter 4), and applying the results of chapter 2, one can assign scale specific weights to each environmental parameter using a simple procedure, making surrogacy more efficient.

Using the marginal effects of the different parameter classes (Ter Braak 1986), one can calculate the relative contribution of each class (measured as explained variance) to the overall performance, and apply these relative values (Fig. 5.1) as the weights of the different parameter classes. The efficiency of surrogates for biodiversity in representing target species is still controversial (Rodrigues and Brooks 2007), however they are the most effective tool for designing reserve networks in an attempt to reduce rates of species extinctions and habitat loss.

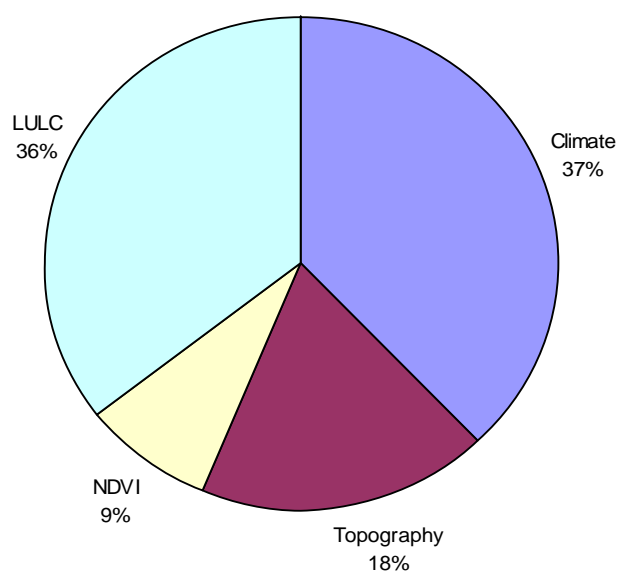


Figure 5.1: Relative importance of the different parameter classes, in percent, at a 100 km² scale

Summary and conclusions

Understanding the determinants of spatial patterns of biodiversity, be it species, communities or ecosystems, is a central question in ecology (Whittaker 1967, MacArthur 1972). Studying the effect of these relationships on biodiversity conservation contributes to conservation scientists and practitioners (Ferrier and Guisan 2006).

The main difficulty in performing large scale analyses on species composition determinants is lack of appropriate data (dedicated biological surveys, resulting in detailed presence-absence or abundance data), especially in the most diverse areas of the world (Elith et al. 2006, Ferrier and Guisan 2006). In chapter 1 I used virtual species to test whether available presence-only (PO) data are indeed

inappropriate for analyzing large scale species composition determinants, as suggested by several studies. The analyses revealed that at large scales, PO data are sufficient for composition analyses, due to excess information in the dataset, resulting from the multitude of species in the analysis. Since the simulations were based on real environmental data, and the PO dataset was created using actual occurrences from the Global Biodiversity Information Facility (GBIF 2008), the results reflect on the true ability of the methods I used to evaluate the relationships between species composition and its environmental determinants. These results allowed me to proceed with the analyses of the relationships between the composition of mammals in the USA and environmental parameters, and how they are affected by spatial scale, described in chapter 2. In that chapter I performed analyses on large scale determinants of mammal species composition in the USA, This kind of analyses were possible based on the results of chapter 1, and provide first evidence of large-scale environmental determinants of species composition, testing long existing theories.

Hybridization of biological and environmental surrogates was first suggested over a decade ago (Kirkpatrick and Brown 1994, Ferrier and Watson 1997), but very few attempts have been made to actually develop such surrogates (e.g. Ferrier and Guisan 2006). One reason for that is the difficulty to combine data from different sources. Biological data are usually in the form of binary data of species in selected survey sites (Legendre et al. 2005), i.e. they are not continuous in space (data exist only for visited sites) or in form (they have a value of 0/1). On the other hand, environmental data are often derived from remote sensing sources, and continuously cover large geographic extents. They also often have continuous values. In order to create spatially continuous biological data, I used an approach developed by Ferrier and Guisan (2006) termed by the authors 'predict first, assemble later', in which the binary presence-absence data for each species is first used to create a predicted distribution map, and then a classification is performed on the stacked distribution maps. Using interchangeable similarity measures i.e. Jackard and Gower (Dunn and Everitt 1982), on the

biological and environmental data respectively, it was possible to use both types of data in a single surrogacy analysis. Chapters 3 and 4, relying on the assumption that the relationships between one group of species and its physical environment can be used to infer on the distribution of other groups, present two novelties – first I introduce a new approach to the evaluation of surrogacy efficiency, which makes surrogacy more applicable from a practitioner's point of view. Second I introduce the first fully hybridized bio-environmental surrogates.

REFERENCES

- Algar, A. C., J. T. Kerr, and D. J. Currie. 2007. A test of Metabolic theory as the mechanism underlying broad-scale species-richness gradients. *Global Ecology and Biogeography* **16**:170-178.
- Allen, A. P., J. H. Brown, and J. F. Gillooly. 2002. Global biodiversity, biochemical kinetics, and the Energetics-Equivalence Rule. *Science* **297**:1545-1548.
- Allen, T. and T. Hoekstra. 1992. *Toward a Unified Ecology*. New York: Columbia UP.
- Araujo, M. B. and C. J. Humphries. 2001. Would environmental diversity be a good surrogate for species diversity? *Ecography* **24**:103-110.
- Arponen, A., A. Moilanen, and S. Ferrier. 2008. A successful community-level strategy for conservation prioritization. *Journal of Applied Ecology* **45**:1436-1445.
- Bar Massada, A., O. Gabay, A. Perevolotsky, and Y. Carmel. 2008. Quantifying the effect of grazing and shrub-clearing on small scale spatial pattern of vegetation. *Landscape Ecology* **23**:327-339.
- Belbin, L. 1993. Environmental Representativeness - Regional Partitioning and Reserve Selection. *Biological Conservation* **66**:223-230.
- Belbin, L. 1995. A multivariate approach to the selection of biological reserves. *Biodiversity and Conservation* **4**:951-963.
- Bonn, A. and K. J. Gaston. 2005. Capturing biodiversity: selecting priority areas for conservation using different criteria. *Biodiversity and Conservation* **14**:1083-1100.
- Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. *Ecology* **73**:1045-1055.
- Brotons, L., W. Thuiller, M. B. Araujo, and A. H. Hirzel. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* **27**:437-448.

- Carmel, Y. and L. Stoller-Cavari. 2006. Comparing environmental and biological surrogates for biodiversity at a local scale. *Israel Journal of Ecology and Evolution* **52**:11-28.
- Chiarucci, A., F. D'Auria, V. De Dominicis, A. Lagana, C. Perini, and E. Salerni. 2000. Using vascular plants as a surrogate taxon to maximize fungal species richness in reserve design. *Conservation Biology* **19**:1644-1652.
- Clements, F. 1916. *Plant succession: an analysis of the development of vegetation*. Carnegie Institution of Washington.
- Connell, J. H. 1961. The influence of interspecific competition and other factors on the distribution of the barnacle *Chtamalus stellatus*. *Ecology* **42**:710-723.
- Costanza, R., R. D'Arge, R. de Groot, S. Farber, M. Grasso, B. Hannon, K. Limburg, S. Naeem, R. O'Neill, and J. Paruelo. 1997. The value of the world's ecosystem services and natural capital. *Nature* **387**:253-260.
- Crawley, M. J. and J. E. Harral. 2001. Scale dependence in plant biodiversity. *Science* **291**:864-868.
- Cushman, S. A. and K. McGarigal. 2002. Hierarchical, multi-scale decomposition of species-environment relationships. *Landscape Ecology* **17**:637-646.
- Dobson, A. P., J. P. Rodrigues, W. M. Roberts, and D. S. Wilcove. 1997. Geographic distribution of endangered species in the United States. *Science* **275**:553-553.
- Dunn, G. and B. S. Everitt. 1982. *An introduction to mathematical taxonomy*. Cambridge university press, Cambridge.
- Ehrlich, P. R. and E. O. Wilson. 1991. Biodiversity studies: Science and policy. *Science* **253**:758-762.
- Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. Leathwick, A. Lhmann, J. Li, L. G. Lohmann, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N.

- E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**:129-151.
- ESRI. 1999. ArcView GIS. ESRI.
- Everitt, B. S. 1993. Cluster analysis. Third edition. John Wiley & Sons Inc., New-York.
- Fairbanks, D. H. 2000. Physio-climatic classification of South Africa's woodland biome. *Plant Ecology* **149**:71-89.
- Faith, D. P. and P. A. Walker. 1996. Environmental diversity: the best-possible use of surrogate data for assessing the relative biodiversity of sets of areas. *Biodiversity and Conservation* **5**:399-415.
- Ferrier, S. 2002. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Systematic Biology* **51**:331-363.
- Ferrier, S., M. Drielsma, G. Manion, and G. Watson. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity and Conservation* **11**:2309-2338.
- Ferrier, S., D. P. Faith, A. Arponen, and M. Drielsma. 2009. Community-level approaches to spatial conservation prioritization. Page 320 in A. Moilanen, K. A. Wilson, and H. P. Possingham, editors. *Spatial conservation prioritization*. Oxford University Press, Oxford.
- Ferrier, S. and A. Guisan. 2006. Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology* **43**:393-404.
- Ferrier, S. and G. Watson. 1997. An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity. New South Wales National Parks and Wildlife Service.
- Fleishman, E., T. J. R. Mac Nally, and D. D. Murphy. 2005. Using indicator species to predict species richness of multiple taxonomic groups. *Conservation Biology* **19**:1125-1137.

- Garson, J., A. Aggarwal, and S. Sarkar. 2002a. Birds as surrogates for biodiversity: an analysis of a data set from southern Quebec. *Journal of bioscience* **27**:347-360.
- Garson, J., A. Aggarwal, and S. Sarkar. 2002b. Birds as surrogates for biodiversity: an analysis of a data set from southern Quebec. *Journal of bioscience* **27**:347-360.
- GBIF. 2008. GBIF Training Manual 1: Digitisation of History Collections Data, version 1.0. Global biodiversity information facility, Copenhagen.
- Gitay, H., A. Suarez, R. Watson, and D. Dokken. 2002. Climate change and biodiversity. Intergovernmental Panel on Climate Change.
- Gleason, H. 1939. The individualistic concept of the plant association. *American Midland Naturalist*:92-110.
- Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**:325-338.
- Gower, J. C. 1971. A general coefficient of similarity and some of its properties. *Biometrics* **27**:857-874.
- Graham, C., S. Ferrier, F. Huettman, C. Moritz, and A. Peterson. 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution* **19**:497-503.
- Guisan, A. and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* **135**:147-186.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolates climate surfaces for global land areas. *International journal of climatology* **25**:1965-1978.

- Hijmans, R. J., K. A. Garrett, Z. Huaman, D. P. Zhang, M. Schreuder, and M. Bonierbale. 2000. Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. *Conservation Biology* **14**:1755-1765.
- Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* **83**:2027-2036.
- Hubbell, S. 1997. A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs* **16**:9-21.
- Hubbell, S. 2005. Neutral theory in community ecology and the hypothesis of functional equivalence. *Ecology* **19**:166-172.
- Hutchinson, G. E. 1957. Concluding remarks. *Symposium on Quantitative Biology* **22**:415-427.
- Jones, M., H. Tuomisto, D. Borcard, P. Legendre, D. Clark, and P. Olivas. 2008. Explaining variation in tropical plant community composition: influence of environmental and spatial data quality. *Oecologia* **155**:593-604.
- Kadmon, R. and A. Danin. 1999. Distribution of plant species in Israel in relation to spatial variation in rainfall. *Journal of Vegetation Science* **10**:421-432.
- Kadmon, R., O. Farber, and A. Danin. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications* **14**:401-413.
- Kadmon, R. and J. Heller. 1998. Modelling faunal responses to climatic gradients with GIS: land snails as a case study. *Journal of Biogeography*:527-539.
- Kati, V., P. Devillers, M. Dufrene, A. Legakis, D. Vokou, and P. Lebrun. 2004a. Hotspots, complementarity or representativeness? designing optimal small-scale reserves for biodiversity conservation. *Biological Conservation* **120**:471-480.

- Kati, V., P. Devillers, M. Dufrene, A. Legakis, D. Vokou, and P. Lebrun. 2004b. Testing the value of six taxonomic groups as biodiversity indicators at a local scale. *Conservation Biology* **18**:667-675.
- Kays, R. and D. Wilson. 2002. *Mammals of North America*. Princeton Univ Pr.
- Kerr, J. T., A. Sugar, and L. Packer. 2000. Indicator taxa, rapid biodiversity assessment and nestedness in an endangered ecosystem. *Conservation Biology* **14**:1726-1734.
- Kirkpatrick, J. B. and M. J. Brown. 1994. A comparison of direct and environmental domain approaches to planning reservation of forest higher-plant communities and species in Tasmania. *Conservation Biology* **8**:217-224.
- Knight, A. T., R. M. Cowling, M. Rouget, A. Balmford, A. M. Lombard, and B. M. Campbell. 2008. Knowing but not doing: selecting priority conservation areas and the research-implementation gap. *Conservation Biology* **22**:610-617.
- Lawler, J. J. and D. White. 2008. Assessing the mechanisms behind successful surrogates for biodiversity in conservation planning. *Animal Conservation* **11**:270-280.
- Le Honerou, H. N. 1981. Impact of man and his animals on Mediterranean vegetation. *Ecosystems of the World, II, Mediterranean Shrublands* (eds F. Di Castri, D.W. Goodall & R.L. Specht). Elsevier Scientific, Amsterdam.
- Leathwick, J., J. M. Overton, and M. Mcleod. 2003. An environmental domain classification of New Zealand and its use as a tool for biodiversity management. *Conservation Biology* **17**:1612-1623.
- Legendre, P., D. Borcard, and P. Peres-Neto. 2005. Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs* **75**:435-450.
- Legendre, P. and L. Legendre. 1998. *Numerical Ecology*. second edition. Elsevier Science, Amsterdam.

- Levin, S. A. 2000. Multiple scales and the maintenance of biodiversity. *Ecosystems* **3**:498-506.
- Loiselle, B. A., P. M. Jorgensen, T. Consiglio, I. Jimenez, J. G. Blake, L. G. Lohmann, and O. M. Montiel. 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography* **35**:105-116.
- MacArthur, R. H. 1972. *Geographical ecology: patterns in the distribution of species*. Harper and Row, New York.
- MacArthur, R. H. and E. O. Wilson. 1967. The theory of island biogeography. *Monographs in Population Biology* **1**.
- MacNally, R., E. Fleishman, J. P. Fay, and D. D. Murphy. 2003. Modelling butterfly species richness using mesoscale environmental variables: model construction and validation for mountain ranges in the Great Basin of western North America. *Biological Conservation*:21-31.
- Maddock, A. and M. A. Du Plessis. 1999. Can species data only be appropriately used to conserve biodiversity? *Biodiversity and Conservation* **8**:603-615.
- Manly, B. F. J. 2007. *Randomization, bootstrap and Mote Carlo methods in biology*. third edition. Taylor and Francis Group, Boca Raton.
- Margules, C. R. and R. L. Pressey. 2000. Systematic conservation planning. *Nature* **405**:243-253.
- Margules, C. R., R. L. Pressey, and P. H. Williams. 2002. Representing biodiversity: data and procedures for identifying priority areas for conservation. *Journal of Bioscience* **27**:309-326.
- Margules, C. R. and J. L. Stein. 1989. Patterns in the distributions of species and the selection of nature reserves: An example from Eucalyptus forests in South-eastern New South Wales. *Biological Conservation* **50**:219-238.
- May, R. 1984. An overview: real and apparent patterns in community structure. *Ecological communities: conceptual issues and the evidence*. Princeton University Press, Princeton, New Jersey, USA:3-16.

- McGarigal, K., S. A. Cushman, M. C. Neel, and E. Ene. 2002. FRAGSTAT. Pages Spatial Pattern Analysis Program for Categorical Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: www.umass.edu/landeco/research/fragstats/fragstats.html.
- Menge, B. and A. Olson. 1990. Role of scale and environmental factors in regulation of community structure. *Trends in Ecology & Evolution* **5**:52-57.
- Minchin, P. 1987. Simulation of multidimensional community patterns: towards a comprehensive model. *Plant Ecology* **71**:145-156.
- Naveh, Z. 1975. The evolutionary significance of fire in the Mediterranean region. *Vegetatio* **29**:199-208.
- Naveh, Z. and J. Dan. 1973a. The human degradation of Mediterranean landscapes in Israel. In: di Castri, F., Mooney, H.A. (Eds.), *Mediterranean-Type Ecosystems, Origin and Structure*. Springer-Verlag, New York.
- Naveh, Z. and J. Dan. 1973b. The human degradation of Mediterranean landscapes in Israel. *in* F. di Castri and H. A. Mooney, editors. *Mediterranean-type ecosystems, origin and structure*. Springer-Verlag, New-York.
- Naveh, Z. and P. Kutiel. 1986. Changes in the Mediterranean vegetation of Israel in response to human habitation and land use. Pages 259-296 *in* G. M. Woodwell, editor. *The earth in transition, patterns and processes of biotic impoverishment*. Cambridge university press, Cambridge.
- Nogués-Bravo, D., M. Aránjo, T. Romdal, and C. Rahbek. 2008. Scale effects and human impact on the elevational species richness gradients. *Nature* **453**:216-219.
- Noss, R. F. 1987. From plant communities to landscapes in conservation inventories: A look at the nature conservancy (USA). *Biological Conservation* **41**:11-37.

- Noss, R. F. 1990. Indicators for monitoring biodiversity: a hierarchical approach. *Conservation Biology* **4**:355-364.
- Oliver, I., A. Holmes, J. M. Dangerfield, M. Gillings, A. J. Pik, D. R. Britton, M. Holley, M. E. Montgomery, M. Raison, V. Logan, R. L. Pressey, and A. J. Beattie. 2004. Land systems as surrogates for biodiversity in conservation planning. *Ecological Applications* **14**:284-503.
- Orme, C. D. L., R. G. Davies, M. Burgess, F. Eigenbrod, N. Pickup, V. A. Olson, A. J. Webster, T. S. Ding, P. C. Rasmussen, R. S. Ridgely, A. J. Stattersfield, P. M. Bennet, T. M. Blackburn, K. J. Gaston, and I. P. F. Owens. 2005. Global hotspots of species richness are not congruent with endemism or threat. *Nature* **436**:1016-1019.
- Palmer, M. and P. Dixon. 1990. Small scale environmental heterogeneity and the analysis of species distributions along gradients. *Journal of Vegetation Science* **1**:57-65.
- Patrick, R. 1963. The structure of Diatom communities under varying ecological conditions. *Annals of the New York Academy of Sciences* **108**:359-365.
- Perevolotsky, A., Ettinger, E., Schwartz-Tzachor, R., Yonatan, R. 2002. Management of fuel breaks in the Israeli Mediterranean Ecosystem: the case of Ramat Hanadiv Park. *Journal of Mediterranean Ecology* **3**:13-22.
- Pharo, E., A. Beattie, and R. Pressey. 2000. Effectiveness of using vascular plants to select reserves for bryophytes and lichens. *Biological Conservation* **96**:371-378.
- Phillips, S., R. Anderson, and R. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**:231-259.
- Pianka, E. R. 1966. Latitudinal gradients in species diversity: a review of concepts. *The American Naturalist* **100**:33-46.
- Poiani, K. A., B. D. Richter, M. G. Anderson, and H. E. Richter. 2000. Biodiversity conservation at multiple scales: functional sites, landscapes, and networks. *Bioscience* **50**:133-146.

- Ponder, W. F., G. A. Carter, P. Flemons, and R. R. Chapman. 2001. Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology* **15**:648-657.
- Prendergast, J. R., R. M. Quinn, and J. H. Lawton. 1999. The gaps between theory and practice in selecting nature reserves. *Conservation Biology* **13**:484-492.
- Pressey, R. L., C. J. Humphries, C. R. Margules, R. I. Vanewright, and P. H. Williams. 1993. Beyond Opportunism - Key Principles for Systematic Reserve Selection. *Trends in Ecology & Evolution* **8**:124-128.
- Pyron, R. and F. Burbrink. 2009. Can the Tropical Conservatism Hypothesis explain temperate species richness patterns? An inverse latitudinal biodiversity gradient in the New World snake tribe Lampropeltini. *Global Ecology and Biogeography* **18**:406-415.
- Rahbek, C. and G. Graves. 2001. Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences* **98**:4534.
- Reese, G. C., K. R. Wilson, J. Hoeting, and C. H. Flather. 2005. Factors affecting species distribution predictions: a simulation modeling experiment. *Ecological Applications* **15**:554-564.
- Reyers, B., D. H. Fairbanks, A. S. V. Jaarsveld, and M. Thompson. 2001. Priority areas for the conservation of South African vegetation: a coarse-filter approach. *Diversity and Distributions* **7**:79-95.
- Ricketts, T. H., E. Dinerstein, T. Boucher, T. M. Brooks, S. H. M. Butchart, M. Hoffmann, J. F. Lamoreux, J. Morrison, M. Parr, J. D. Pilgrim, A. S. L. Rodrigues, W. Secherst, G. E. Wallace, K. Berlin, J. Bielby, N. D. Burgess, D. R. Church, N. Cox, D. Knox, C. Loucks, G. W. Luck, L. L. Master, R. Moore, R. Naidoo, R. S. Ridgely, G. E. Schartz, G. Shire, H. Strand, W. Wettengel, and E. Wikramanayake. 2005. Pinpointing and preventing imminent extinctions. *Proceedings of the National Academy of Science* **102**:18497-18501.

- Rodrigues, A. S. L. and T. M. Brooks. 2007. Shortcuts for biodiversity conservation planning: the effectiveness of surrogates. *Annu. Rev. Ecol. Sys.* **38**:713-737.
- Rodriguez, J., J. Hortal, and M. Nieto. 2006. An evaluation of the influence of environment and biogeography on community structure: the case of Holarctic mammals. *Journal of Biogeography* **33**:291-303.
- Rosenzweig, M. L. 1995. *Species diversity in space and time*. Cambridge University Press, Cambridge.
- Sarkar, S., R. L. Pressey, D. P. Faith, C. R. Margules, T. Fuller, D. M. Stoms, A. Moffet, K. A. Wilson, K. J. Williams, P. H. Williams, and S. J. Andelman. 2006. Biodiversity conservation planning tools: present status and challenges for the future. *Annual review of environment and resources* **31**:123-159.
- Sastre, P. and J. M. Lobo. 2009. Taxonomic survey biases and the unveiling of biodiversity patterns. *Biological Conservation* **142**:462-467.
- Sauberer, N., K. Zulka, M. Abensperg-Traun, H. Berg, G. Bieringer, N. Milasowszky, D. Moser, C. Plutzer, M. Pollheimer, and C. Storch. 2004. Surrogate taxa for biodiversity in agricultural landscapes of eastern Austria. *Biological Conservation* **117**:181-190.
- Schwartz, M. W. 1999. Choosing the appropriate scale of reserves for conservation. *Annual Review of Ecology and Systematics* **30**:83-103.
- Shmida, A. and M. Wilson. 1985. Biological determinants of species diversity. *Journal of Biogeography*:1-20.
- Shoshany, M. 2000. Satellite remote sensing of natural Mediterranean vegetation: a review within an ecological context. *Progress in physical geography* **24**:153-178.
- Shriner, S. A., K. R. Wilson, and C. H. Flather. 2006. Reserve networks based on richness hotspots and representation vary with scale. *Ecological Applications* **16**:1660-1673.

- Stoms, D. M. 2000. GAP management status and regional indicators of threats to biodiversity. *Landscape Ecology* **15**:21-33.
- Svenning, J. and F. Skov. 2005. The relative roles of environment and history as controls of tree species composition and richness in Europe. *Journal of Biogeography* **32**:1019-1033.
- Ter Braak, C. J. F. 1986. Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67**:1167-1179.
- Ter Braak, C. J. F. and P. Smilauer. 2002. CANOCO reference manual and CanoDraw for Windows user's guide: software for canonical community ordination (version 4.5). Page 500. Microcomputer power, Ithaca, New-York.
- Ter Braak, C. J. F. and P. F. M. Verdonschot. 1995. Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Sciences* **57**:255-289.
- Trakhtenbrot, A. and R. Kadmon. 2005. Environmental cluster analysis as a tool for selecting complementary networks of conservation sites. *Ecological Applications* **15**:335-345.
- Trakhtenbrot, A. and R. Kadmon. 2006. Effectiveness of environmental cluster analysis in representing regional species diversity. *Conservation Biology* **20**:1087-1098.
- Tsoar, A., O. Allouche, O. Steinitz, D. Rotem, and R. Kadmon. 2007. A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions* **13**:397-405.
- Vandermeer, J. H. 1972. Niche theory. *Annual Review of Ecology and Systematics* **3**:107-132.
- Virolainen, K. M., P. Ahlroth, E. Hyvarinen, E. Korkeamaki, J. Mattila, J. Paivinen, T. Rintala, T. Suomi, and J. Suhonen. 2000. Hot spots, indicator taxa, complementarity and optimal networks of taiga. *Proceedings of the royal society of London* **267**:1143-1147.
- Wagner, H. H. and M. J. Fortin. 2005. Spatial analysis of landscapes: concepts and statistics. *Ecology* **86**:1975-1987.

- Ward, H. J. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* **58**:236-244.
- Whittaker, R. H. 1956. Vegetation of the Great Smoky Mountains. *Ecological Monographs* **26**:2-80.
- Whittaker, R. H. 1967. Gradient analysis of vegetation. *Biological Reviews of the Cambridge Philosophical Society* **42**:207-264.
- Whittaker, R. J., K. J. Willis, and R. Field. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography* **28**:453-470.
- Wiens, J. 1989. Spatial scaling in ecology. *Functional Ecology* **3**:385-397.
- Willig, M., D. Kaufman, and R. Stevens. 2003. Latitudinal gradients of biodiversity: Pattern, process, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics* **34**:273-309.
- Willis, K. J. and R. J. Whittaker. 2002. Species diversity - scale matters. *Science* **295**:1245-1248.
- Wilson, R. J., C. D. Thomas, R. Fox, D. B. Roy, and W. E. Kunin. 2004. Spatial patterns in species distributions reveal biodiversity change. *Nature* **432**:393-396.
- Wilson, W. G., P. Lundberg, D. P. Vasquez, J. B. Shurin, M. D. Smith, W. Langford, K. L. Gross, and G. G. Mittlebach. 2003. Biodiversity and species interactions: extending Lotka-Volterra community theory. *Ecology Letters* **6**:944-952.
- Worm, B., E. Barbier, N. Beaumont, J. Duffy, C. Folke, B. Halpern, J. Jackson, H. Lotze, F. Micheli, and S. Palumbi. 2006. Impacts of biodiversity loss on ocean ecosystem services. *Science* **314**:787.
- Yom-Tov, Y. and R. Kadmon. 1998. Analysis of the distribution of insectivorous bats in Israel. *Diversity and Distributions* **4**:63-70.

Appendix 1 – A list of woody and geophyte plant species in the Carmel surrogate analyses

<u>Scientific name</u>	<u>Common name</u>
<u>Woody plant species</u>	
<i>Arbutus andrachne</i>	Greek strawberry tree
<i>Asparagus aphyllus</i>	Prickly asparagus
<i>Calicotome villosa</i>	Spiny broom
<i>Cistus creticus</i>	Sage leaved rock rose
<i>Cistus salviiifolius</i>	Soft-hairy rock rose
<i>Clematis cirrhosa</i>	Virg in's Bower
<i>Ephedra foemina</i>	Leafless ephedra
<i>Fumana arabica</i>	Arabian cistus
<i>Fumana thymifolia</i>	Thyme rock rose
<i>Genista fasselata</i>	
<i>Laurus nobilis</i>	True Laurel
<i>Lonicera etrusca</i>	Italian honeysuckle
<i>Majorana syriaca</i>	Wild Marjoram
<i>Micromer fruticosa</i>	White leaved savory
<i>Myrtus communis</i>	common myrtle
<i>Olea europaea</i>	Olive tree
<i>Osyris alba</i>	Poet's Cassia
<i>Phagnalon rupestre</i>	African fleabane
<i>Phillyrea latifolia</i>	Broad-leaved phillyrea
<i>Phlomis viscosa</i>	Shrubby Jerusalem sage

<i>Pinus halepensis</i>	Aleppo pine
<i>Pinus pinea</i>	Stone pine
<i>Pistacia lentiscus</i>	Mastic tree
<i>Pistacia palaestina</i>	Terebinth tree
<i>Prasium majus</i>	Great hedge-nettle
<i>Quercus calliprinos</i>	Kermes Oak
<i>Rhamnus lycioides</i>	Palestine buckthorn
<i>Rhamnus alaternus</i>	Italian buckthorn
<i>Rubia tenuifolia</i>	Common Madder
<i>Ruscus aculeatus</i>	
<i>Salvia fruticosa</i>	Three-lobed sage
<i>Sarcopoter spinosum</i>	Prickly burnet
<i>Satureja thymbra</i>	Savory of Crete
<i>Stachys palaestina</i>	
<i>Thamua sp.</i>	
<i>Theucrium divaricatum</i>	
<i>Thymelaea hirsute</i>	Shaggy sparrow-wort

Geophyte species

<i>Allium trifoliatum</i>	
<i>Anemone coronaria</i>	Crown anemone
<i>Arisarum vulgare</i>	Aha'rons rod
<i>Arum dioscoridis</i>	Spotted arum
<i>Asphodelus ramosus</i>	Common asphodel
<i>Ruscus aculeatus</i>	

<i>Emex spinosa</i>	Spiny dock
<i>Scila hyacinthoides</i>	Hyacinth squill
<i>Ophrys tricolor</i>	
<i>Ophrys fleischmanii</i>	
<i>Gynandriris sisyrinchium</i>	Barbary nut
<i>Narcissus tazetta</i>	Common narcissus
<i>Ornithogalum narbonense</i>	Narbonne star-of-Bethlehem
<i>Pancratium parviflorum</i>	Small flowered Pancratium
<i>Cyclamen persicum</i>	Persian Cyclamen
<i>Cephalanthera longifolia</i>	Long-leaf Helleborine
<i>Orchis galilaea</i>	Orchid
<i>Limodorum abortivum</i>	Violet Limodore
<i>Allium carmeli</i>	
<i>Allium neapolitanum</i>	Naples garlic
<i>Serapias vomeracea</i>	Snake tongue orchid
<i>Umbilicus intermedius</i>	Common pennywort
<i>Gagea commutate</i>	Stolonous Gagea

יחסים בין משתני סביבה להרכב מינים והשלכותיהם על סמנים של
מגוון ביולוגי - מחקר בסקאלות מרובות

רפי קנט

יחסים בין משתני סביבה להרכב מינים והשלכותיהם על סמנים של
מגוון ביולוגי - מחקר בסקאלות מרובות

לצורך מילוי חלקי של הדרישות לקבלת התואר דוקטור לפילוסופיה

רפי קנט

הוגש לסנט הטכניון – מכון טכנולוגי לישראל

דצמבר 2010

חיפה

טבת ה'תשע"א

המחקר נעשה בהנחיית פרופ"ח יוחאי כרמל

אני מודה לטכניון על התמיכה הכספית הנדיבה

תקציר

היחסים בין מינים לסביבתם, והגורמים להרכב המינים במרחב הם נושאים מרכזיים באקולוגיה. הבנת היחסים הללו עשויה לתרום לשמירה על המגוון הביולוגי. קיימות מספר תיאוריות המנסות להסביר תופעות בסקאלות גדולות, הקשורות בעושר המינים. לדוגמא הגראדיינט במספר המינים מקו המשווה לכיוון הקטבים, מוסבר על-ידי תיאוריות מתחרות כגון יחס עושר מינים לשטח, הלקוח מביו-גיאוגרפיה של איים, כמות האנרגיה הזמינה, וקצבי הכחדות והתמיינות שונים לאורך הגראדיינט. לגבי הרכב מינים קיימות פחות תיאוריות, אולם קיימים שלושה מודלים עקרוניים. המודל הראשון הוא ניטרלי, ולפיו כל המינים שווים זה לזה מבחינה אקולוגית ותחרותית, וההבדלים בהרכב המינים הם תוצאה של תהליכי אקראי של הפצה והתבססות. המודל השני הוא המודל הסביבתי, לפיו הרכב המינים בשטח מסוים נקבע על-ידי מכלול התנאים הסביבתיים באותו שטח. לפי המודל השלישי הרכב המינים הוא תוצאה של אינטראקציות בין המינים השונים, בתוך ובין רמות טרופיות שונות (למשל תחרות וטריפה בהתאמה). לצורך ניתוח היחסים בין מינים לסביבתם, ולצורך תכנון לשמירת טבע בשטחים נרחבים יש צורך בנתוני תפוצה של מינים רבים, ובאזורים נרחבים, אשר אינם קיימים. ההשקעה הדרושה כדי לאסוף אותם הופכת אותם לבלתי-אפשריים לאיסוף. עבודה זו מנסה לקדם את הכלים המדעיים לשמירה על המגוון הביולוגי בשני מישורים, ולתת פיתרון, לפחות חלקי, לבעיית זמינות הנתונים. ראשית ביצעתי השוואה בין שיטות להפקת מפות של סמנים ביולוגיים וסביבתיים למגוון ביולוגי (surrogates for biodiversity), והצעתי מדד נוסף לבחינת היעילות של סמנים כאלו, המבוסס על מידת השווינויות בגודל של היחידות השונות המרכיבות את מפות הסמנים. כיום קיימים שני סוגי סמנים למגוון ביולוגי. סמנים סביבתיים מבוססים על חלוקה של המרחב ליחידות בעלות תנאי-סביבה דומים, וההנחה שעומדת בבסיס השיטה היא שייצוג של כלל תנאי הסביבה בשטח מסוים יביא לייצוג של כלל המגוון הביולוגי באותו שטח, לפי המודל הסביבתי להרכב מינים. סמנים ביולוגיים הם קבוצת מינים, שיש עליהם נתוני תפוצה טובים וניתן להסיק מהתפוצה שלהם לגבי התפוצה של כלל המגוון. יש תיעוד של שימוש במספר רב של קבוצות מינים כסמנים, לדוגמא דו-חיים, פרפרים, מיני צומח מעוצה וציפורים. אני הכנתי מפות של סמנים משני הסוגים, כלומר ביולוגיים וסביבתיים, עבור הכרמל. לצורך כך השתמשתי בסט קיים של נתוני תפוצה של מיני גיאופיטים ומיני צומח מעוצה שנסקרו ב-100 מיקומים אקראיים בכרמל ליצירת סמנים ביולוגיים. הנתונים האלו שימשו ליצירת מפות תפוצה ל-23 מיני גיאופיטים ו-37 מיני צומח מעוצה על-ידי מודלים של רגרסיה לוגיסטית. לצורך יצירת סמנים סביבתיים השתמשתי בנתוני אקלים, טופוגרפיה, יצרנות ראשונית ומדד של הפרעה אנושית. את הסמנים יצרתי בחמש שיטות סיווג שונות של ניתוח-אשכולות, ארבע שיטות הירארכיות ואחת לא

הירארכית, והשוויתי את היעילות של הסמנים כתלות בשיטת הסיווג. השתמשתי בשני מדדים לבדיקת היעילות של הסמנים. הראשון הוא מספר מיני המטרה המיוצגים על-ידי ייצוג כל היחידות של הסמנים. המדד השני מוצג כאן בפעם הראשונה כמדד ליעילות של סמנים למגוון ביולוגי, והוא כולל חישוב של השוויוניות בגודל היחידות השונות של הסמנים. ההשוואה העלתה ששיטת הסיווג היעילה ביותר ליצירת סמנים, בין אם ביולוגיים או סביבתיים, היא שיטת וורד למינימום שונות (Ward's method of minimum variance). בנוסף פיתחתי סוג חדש של סמנים, המבוססים על שילוב בין סוגי הסמנים הקיימים, כלומר סמנים ביו-סביבתיים. כבר בשנות ה-90 של המאה הקודמת הציעו חוקרים ששילוב של שני סוגי הסמנים, ושימוש בשני סוגי המידע (ביולוגי וסביבתי) יוביל ליכולת ייצוג טובה יותר של המגוון הביולוגי על-ידי שימוש בסמנים, אולם זאת הפעם הראשונה שסמנים כאלו מפותחים ונבדקים בצורה כמותית. בעבודה זאת מוצגות שלוש גישות ליצירת הסמנים היברידיים למגוון ביולוגי. שתי גישות היררכיות דו-שלביות מבוססות על עיקרון הנישה האקולוגית הממומשת. הגישה הראשונה – גישת הנישה, כוללת בשלב הראשון חלוקה של המרחב ליחידות בעלות תנאי סביבה דומים, בדומה לסמנים סביבתיים, ובשלב השני חלוקת משנה של כל יחידה לתת-יחידות בעלות הרכב מינים דומה. הגישה השנייה – גישת הנישה ההפוכה מתבצעת בדיוק באותה השיטה, אך בסדר הפוך, כלומר ראשית מחלקים את המרחב ליחידות בעלות הרכב מינים דומה, ולאחר מכן חלוקת משנה של כל יחידה לתת-יחידות בעלות תנאים סביבתיים דומים. הגישה השלישית היא לא היררכית וכוללת הכנה של שתי מפות סמנים, סביבתיים וביולוגיים בנפרד ואז חיתוך של המפות אחת עם השנייה לקבלת יחידות חדשות. בבדיקה שערכתי, הסמנים החדשים ייצגו את מיני המטרה בצורה כמעט מושלמת, ללא תלות בגישה בה השתמשתי ליצירת הסמנים הביו-סביבתיים. המטרה העיקרית השנייה של עבודה זאת הייתה לכמת את מידת השינוי ביחסים בין הרכב מינים ותנאי הסביבה המשפיעים על הרכב זה, כתלות בסקאלה המרחבית. המכשול העיקרי לביצוע ניתוחים מסוג זה הוא זמינות של נתוני תפוצה של מינים. בסקאלות מרחביות גדולות, הנתונים היחידים שקיימים הם נתוני אוספים, שמכילים מידע מסוג נוכחות-בלבד. הבעיה עם נתוני נוכחות-בלבד היא כפולה. ראשית הנתונים אינם מבוססים על סקרים ייעודיים ולכן אינם מכסים את שטח המחקר בצורה אחידה או אקראית, כמו נתוני נוכחות-היעדרות. נתונים מסוג זה מכילים הטיה משלושה סוגים עיקריים. הטיה גיאוגרפית, הנובעת באופן ישיר מהפיזור של הנתונים במרחב, בדרך-כלל קיימים נתונים רבים בקרבה רבה לכבישים ואזורים מיושבים ולעומת זאת יש מיעוט נתונים באזורים שהגישה אליהם היא יותר קשה. הטיה נוספת היא סביבתית, כלומר לא כל תנאי הסביבה מיוצגים בנתונים. הסוג האחרון של הטיה הוא טקסונומי, כלומר ישנם מינים פופולריים יותר או בולטים יותר שיהיו מיוצגים בעודף בנתונים, ולעומתם מינים פופולריים פחות או

קריפטיים יותר יהיו מיוצגים בחסר. בנוסף, אי-הימצאות של מין מסוים במיקום כלשהוא אינה מעידה על היעדרות אמיתית, אלא על כך שאיש לא תיעד את אותו המין באותו המיקום. מסיבות אלו נתונים כאלו נחשבים כלא מתאימים לבחינת היחסים בין הרכב מינים לסביבה. כן קיימים מודלים של בתי-גידול עבור מין בודד המבוססים על נתוני נוכחות בלבד. הדיון האם ניתן להשתמש בנתונים אלו עבור מודלים נמשך, והדעות עדיין חלוקות לגבי היעילות של מודלים כאלו לחזות תפוצה של מינים. על מנת לבדוק האם ניתן להשתמש בנתוני נוכחות-בלבד, כמו אלו שקיימים במאגרי מידע ממוחשבים גדולים כדוגמת Global Biodiversity Information Facility (GBIF), לצורך ניתוח כמותי של יחסי הרכב מינים ותנאים סביבתיים, ביצעתי סימולציות עם מינים וירטואליים המבוססים על נתוני תצפיות אמיתיות על מינים בארה"ב מתוך GBIF, ועל תנאים סביבתיים אמיתיים מארה"ב. הסימולציות הראו כי לסוג הנתונים לא הייתה השפעה על תוצאות של ניתוח מרובה משתנים של הפרמטרים הקובעים את הרכב המינים של 50 מינים וירטואליים שהיו מפוזרים בשטח היבשתי של ארה"ב. תוצאות הסימולציה אפשרו לי לבצע את השלב האחרון בעבודה זאת, כלומר ניתוח השינוי ביחסי מינים-סביבה של יונקים יבשתיים בשטח היבשתי של ארה"ב, כפונקציה של סקאלה מרחבית. לשם כך חישבתי את אחוזי השונות בהרכב המינים המוסברת על-ידי כל אחת מארבע קבוצות של פרמטרים סביבתיים – אקלים, טופוגרפיה, שימושי-קרקע ותכסית ויצרנות ראשונית. הניתוחים בוצעו בארבע סקאלות מרחביות, עם תאי גריד בגודל שנע בין 100 ל-10,000 קמ"ר. התוצאות של הניתוחים הללו הראו שמשתני אקלים, כגון טמפרטורה ומשקעים, הם הפרמטרים החשובים ביותר בקביעת הרכב מיני היונקים בארה"ב, בכל הסקאלות המרחביות שנבדקו. משתני שימושי קרקע ותכסית היו חשובים גם כן בסקאלות קטנות וסקאלות הביניים, אולם היו פחות חשובים בסקאלה הגדולה ביותר. טופוגרפיה ויצרנות ראשונית היו בעלי השפעה חלשה יותר בסקאלות שנבדקו, והשפעתם נחלשה ככל שהסקאלה המרחבית גדלה. ניתן לשלב בין שני כיווני המחקר שתוארו לעיל, לצורך שיפור של יכולת החיזוי של תפוצת המגוון הביולוגי, ולהשתמש ביכולות אלו לייעול התכנון של שמורות טבע ואזורים מוגנים, בסקאלות יחסית גדולות.